

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model.

Run on: January 8, 2004, 10:25:26 ; Search time 48 Seconds

(without alignments)
2331.297 Million cell updates/sec

Title: US-09-907-907A-42

Perfect score: 3557

Sequence: 1 DGPFLLPRRRALTLQVRA.....TAVLLHNTQLDNERLILPL 705

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

A.GeneSeq_190Jn03:*

1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*

2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*

3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*

4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*

5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*

6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*

7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*

8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*

9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*

10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*

11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*

12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*

13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*

14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*

15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*

16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*

17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*

18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*

19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*

20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*

21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*

22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*

23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*

24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3557	100.0	705	21	AA080732
2	3498	98.3	783	23	ABP69432
3	3030	85.2	675	23	ABG30875
4	2285	64.2	504	22	AA92684
5	1982	55.7	899	22	ABG08545
6	1982	55.7	899	22	ABG17275
7	1917	53.9	439	22	ABG17276
8	1842.5	51.8	748	22	AB58546
9	1228.5	34.5	541	22	ABG08547

10	1191.5	33.5	696	22	AAU33947
11	1191.5	33.5	698	22	AAU36732
12	1165.5	32.8	706	23	ABP40063
13	1158	32.6	721	22	AAU38175
14	1156	32.5	1034	20	AAU3792
15	1155	32.5	734	22	AAU34719
16	1134	31.9	709	22	AAU35396
17	1128.5	31.7	704	22	AAU35289
18	1127	31.7	702	22	AAU3479
19	1127	31.7	723	23	ABP48134
20	1126	31.7	716	23	ABP54549
21	1117	31.4	711	24	ABP78218
22	1111	31.2	706	23	AAU72928
23	1101.5	31.0	684	21	ABP44567
24	1101.5	31.0	684	23	ABP54519
25	1097	30.8	613	22	AAU36162
26	1097	30.8	701	22	AAU36457
27	1086	30.5	709	23	ABP30163
28	1086	30.5	713	23	ABP25678
29	1072	30.1	710	23	ABP25679
30	1063.5	29.9	773	23	ABP55204
31	1054	29.6	737	21	AAU99614
32	1054	29.6	737	22	AAU38106
33	1054	29.6	737	24	ABU00963
34	1045	29.4	752	22	AAU46670
35	1044.5	29.4	913	23	ABP66191
36	1014.5	28.5	753	22	AAU91911
37	997.5	28.0	752	22	AAU81201
38	884	24.9	568	20	AAU35655
39	831.5	23.4	688	19	AAW98709
40	831.5	23.4	688	22	AAU35823
41	829.5	23.3	688	22	AAU35983
42	710	20.0	488	21	AAU44542
43	710	20.0	488	23	ABP54494
44	708	19.9	164	21	ABP58873
45	636	17.9	358	21	AAU44590

ALIGNMENTS

RESULT 1

ID AAB08732

AA080732

02-JAN-2001 (first entry)

AA080732;

Amino acid sequence of a human OLD-35 polypeptide.

OLD-35; OLD-64; OLD-137; OLD-142; OLD-175; cancerous phenotype; cellular senescence; terminal differentiation; growth suppression; aging process; type I interferon; cancer cell; tissue regeneration; ss.

OS Homo sapiens.

FN WO200046391-A2.

XX

PD 10-AUG-2000.

XX

PF 02-FEB-2000; 2000WO-US02920.

XX

PR 02-FEB-1999; 99US-0243277.

XX

(UYCO) UNIV COLUMBIA NEW YORK.

PA

XX

Fisher PB, Leszczynska M;

PI

XX

WPI: 2000-532905/48.

XX

N-PSDB; AAA64608.

DR

XX

Novel isolated nucleic acid encoding an OLD-35 or OLD-64 protein useful

PT in the treatment and detection of e.g. cancer and diseases involving
 PT cellular senescence -
 PS Disclosure; Fig 9B; 115pp; English.
 XX The specification describes OLD-35, OLD-64, OLD-137, OLD-139, OLD-142
 CC and OLD-175 proteins. The Old nucleic acids are useful for reversing
 CC the cancerous phenotype of a cancer cell, determining if a cell is
 CC senescent, growth arrested or terminally differentiated. They are also
 CC useful for reversing the aging process in a cell and degrading specific
 CC RNA in a cell. The genes may also be used as a diagnostic indicator of
 CC cellular senescence, terminal differentiation and/or growth suppression
 CC and as a marker to identify drugs or small molecules that will induce
 CC or inhibit cellular senescence or terminal differentiation and type I
 CC interferons. The combination of Old-35 with other interacting proteins
 CC is useful for targeting the differentiation of specific cells. Old-35
 CC can be used to selectively stabilize specific mRNAs containing adenoviral
 CC rich 3' UTRs. The Old proteins are useful for reversing the cancerous
 CC phenotype of a cancer cell and inhibiting the growth of a cancer cell.
 CC They are also useful for regenerating tissue. The present sequence
 CC represents an Old-35 polypeptide.
 XX
 SQ Sequence 705 AA;
 Query Match 100.0%; Score 3557; DB 21; Length 705;
 Best Local Similarity 100.0%; Pred. No. 1.9e-304;
 Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 DGPFLPRRRALQLOVRAIMSSAGSRAVAVDGNRLKLEISGKLAAPAGSAVVGSD 60
 1 DGPFLPRRRALQLOVRAIMSSAGSRAVAVDGNRLKLEISGKLAAPAGSAVVGSD 60
 61 TAAVAVTAVSKTSPSPQMPVVDYRQKAAAGRIPTVYLREVGTSKXELITSRIIDS 120
 61 TAAVAVTAVSKTSPSPQMPVVDYRQKAAAGRIPTVYLREVGTSKXELITSRIIDS 120
 61 TAAVAVTAVSKTSPSPQMPVVDYRQKAAAGRIPTVYLREVGTSKXELITSRIIDS 120
 121 IRPLFPAGFYDQVLCNLAVDGVNEBDVALINGASVALSLDIPMNGPVGAVRIGIID 180
 121 IRPLFPAGFYDQVLCNLAVDGVNEBDVALINGASVALSLDIPMNGPVGAVRIGIID 180
 121 IRPLFPAGFYDQVLCNLAVDGVNEBDVALINGASVALSLDIPMNGPVGAVRIGIID 180
 181 GEYVNPFRKEMSSSTLNLVVAAGAPKSGIWMLESAENIILOODFCHAKVCKYTOQIIQ 240
 181 GEYVNPFRKEMSSSTLNLVVAAGAPKSGIWMLESAENIILOODFCHAKVCKYTOQIIQ 240
 181 GEYVNPFRKEMSSSTLNLVVAAGAPKSGIWMLESAENIILOODFCHAKVCKYTOQIIQ 240
 181 GEYVNPFRKEMSSSTLNLVVAAGAPKSGIWMLESAENIILOODFCHAKVCKYTOQIIQ 240
 241 GIQOLVKEITGYTKRTPPKLFTSPSPVYTKHLMERYAFVFTDEHDKVSRDEAVNKIR 300
 241 GIQOLVKEITGYTKRTPPKLFTSPSPVYTKHLMERYAFVFTDEHDKVSRDEAVNKIR 300
 241 GIQOLVKEITGYTKRTPPKLFTSPSPVYTKHLMERYAFVFTDEHDKVSRDEAVNKIR 300
 241 GIQOLVKEITGYTKRTPPKLFTSPSPVYTKHLMERYAFVFTDEHDKVSRDEAVNKIR 300
 301 LDTEBQLEKEPPEADPYEIIISFNVVAKVERRSIVLMEYKCDGRLTSLRNVSCVDVF 360
 301 LDTEBQLEKEPPEADPYEIIISFNVVAKVERRSIVLMEYKCDGRLTSLRNVSCVDVF 360
 301 LDTEBQLEKEPPEADPYEIIISFNVVAKVERRSIVLMEYKCDGRLTSLRNVSCVDVF 360
 301 LDTEBQLEKEPPEADPYEIIISFNVVAKVERRSIVLMEYKCDGRLTSLRNVSCVDVF 360
 361 KTLHGSALPQGGQVQVCTVTPDSIESGKSDQVITAINIKDKNFMHYEPPIATNEI 420
 361 KTLHGSALPQGGQVQVCTVTPDSIESGKSDQVITAINIKDKNFMHYEPPIATNEI 420
 361 KTLHGSALPQGGQVQVCTVTPDSIESGKSDQVITAINIKDKNFMHYEPPIATNEI 420
 361 KTLHGSALPQGGQVQVCTVTPDSIESGKSDQVITAINIKDKNFMHYEPPIATNEI 420
 421 GKTVGNRRELGHGALAEKALYVPRPFPPTIRTSVLESNSSSSMAACGSLALMD 480
 421 GKTVGNRRELGHGALAEKALYVPRPFPPTIRTSVLESNSSSSMAACGSLALMD 480
 421 GKTVGNRRELGHGALAEKALYVPRPFPPTIRTSVLESNSSSSMAACGSLALMD 480
 421 GKTVGNRRELGHGALAEKALYVPRPFPPTIRTSVLESNSSSSMAACGSLALMD 480
 481 SGVVISAVAGVAGLVTIKDPEKGEIEDYRLTDIIGIEDYNGDMFKIAGTKGITAL 540
 481 SGVVISAVAGVAGLVTIKDPEKGEIEDYRLTDIIGIEDYNGDMFKIAGTKGITAL 540
 481 SGVVISAVAGVAGLVTIKDPEKGEIEDYRLTDIIGIEDYNGDMFKIAGTKGITAL 540
 481 SGVVISAVAGVAGLVTIKDPEKGEIEDYRLTDIIGIEDYNGDMFKIAGTKGITAL 540
 541 QADIKLPEIPIKIYWEAIQASVAKKEILQIMNKTIKSPRASREKNGEVETVQVPSKR 600
 541 QADIKLPEIPIKIYWEAIQASVAKKEILQIMNKTIKSPRASREKNGEVETVQVPSKR 600
 541 QADIKLPEIPIKIYWEAIQASVAKKEILQIMNKTIKSPRASREKNGEVETVQVPSKR 600
 541 QADIKLPEIPIKIYWEAIQASVAKKEILQIMNKTIKSPRASREKNGEVETVQVPSKR 600
 601 AKFVPGPGVYNKXLOAETGVITISQVDETFSPVAPTPVMEADPTEICKDOEOOLE 660
 601 AKFVPGPGVYNKXLOAETGVITISQVDETFSPVAPTPVMEADPTEICKDOEOOLE 660
 601 AKFVPGPGVYNKXLOAETGVITISQVDETFSPVAPTPVMEADPTEICKDOEOOLE 660
 601 AKFVPGPGVYNKXLOAETGVITISQVDETFSPVAPTPVMEADPTEICKDOEOOLE 660
 661 FGAVYATITTEIRDTGVWVKLYPMTAVLANTQLDNERLNLIP 705

DB 661 FGAVYATITTEIRDTGVWVKLYPMTAVLANTQLDNERLNLIP 705
 RESULT 2
 ABP69432
 ID ABP69432 standard; Protein; 783 AA.
 XX
 AC ABP69432;
 XX
 DT 20-JAN-2003 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 1479.
 XX
 DE Human; genome mapping; gene therapy; food supplement; virus; fungus;
 KW cell-proliferative disorder; neurodegenerative disease; bacterial;
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
 KW arthritis; cytotoxic; immunomodulator; nootropic; neuroprotective;
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
 KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;
 KW antiarthritic.
 KW
 XX
 OS Homo sapiens.
 XX
 XX MO200270539-A2.
 XX
 XX 12-SEP-2002.
 XX
 XX 05-MAR-2002; 2002MO-US05095.
 XX
 XX 05-MAR-2001; 2001US-0799451.
 XX
 XX (HXS-) HXSQ INC.
 XX
 P1 Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
 P1 Xue W, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
 P1 Wehrman T, Wang J, Wang D, Drmanac RT;
 XX
 DR WPI: 2002-759812/82.
 DR
 DR N-PSDB; AB211649.
 XX
 PT New polynucleotides comprising sequences assembled from expressed
 PT sequence tags (ESTs), useful for treating cell-proliferative,
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
 PT platelet or coagulation disorders -
 XX
 PS Claim 9; SEQ ID NO 1479; 1012pp + Sequence Listing; English.
 XX
 PS The invention relates to an isolated polynucleotide (I) comprising a
 CC nucleotide sequence selected from any of 948 sequences
 CC (AB211119-AB212066) or their mature protein coding portion, active domain
 CC coding protein or complementary sequences. The polynucleotides are useful
 CC for identifying expressed genes or for physical mapping of human genome.
 CC The encoded polypeptides (ABP68902-ABP6949) are useful as molecular
 CC weight markers, as a food supplement, for generating antibodies, in
 CC cell-proliferative disorders (cancer), neurodegenerative diseases
 CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
 CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
 CC disorders, platelet or coagulation disorders, wound, burns, incision,
 CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
 CC parasitic, arthritis, etc.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from Wipo
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 783 AA;
 Query Match 98.3%; Score 3498; DB 23; Length 783;
 Best Local Similarity 98.9%; Pred. No. 3.7e-299;
 Matches 692; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 DGPFLPRDRALTOUQVLAWSAGSRAVADLGNRKLEISGKLAFFADSGAVVQSGD 60
 DB 18 DGPFLPRDRALTOUQVLAWSAGSRAVADLGNRKLEISGKLAFFADSGAVVQSGD 77
 QY 61 TAVMTAVAKTSPSPQMPVVDYRQKAAAGRIPTNVLREPVSGDEILTSRIIDS 120
 DB 78 TAVMTAVAKTSPSPQMPVVDYRQKAAAGRIPTNVLREPVSGDEILTSRIIDS 137
 QY 121 IRPLPAGFYDTQVLCNLAVDGVNPDVLAINGASVALSLSDIPMNGPVAGVRIGID 180
 DB 138 IRPLPAGFYDTQVLCNLAVDGVNPDVLAINGASVALSLSDIPMNGPVAGVRIGID 197
 QY 181 GRYVNPFRKEMSSSTLNVLVAGAPRSQIVMLBASANTLLQDFCHAIKVGKXTQIIQ 240
 DB 198 GRYVNPFRKEMSSSTLNVLVAGAPRSQIVMLBASANTLLQDFCHAIKVGKXTQIIQ 257
 QY 241 GYQOLVKEGTVKTRPQKLPSPSPVKTTHKLMERLAVPTDYHDKVSSDEAVNKIR 300
 DB 258 CIOQLVKEGTVKTRPQKLPSPSPVKTTHKLMERLAVPTDYHDKVSSDEAVNKIR 317
 QY 301 LDTEBOLKEKPEADPEYELIESFNVAKEVRSIVLNEYKRCQGRDLTSLANVSCVDWF 360
 DB 318 LDTEBOLKEKPEADPEYELIESFNVAKEVRSIVLNEYKRCQGRDLTSLANVSCVDWF 377
 QY 361 KTLHGSALFQRCQTOVLTCTVFDLSGSKSDQVITAINIGIKDKNFMHYEPYATNMI 420
 DB 378 KTLHGSALFQRCQTOVLTCTVFDLSGSKSDQVITAINIGIKDKNFMHYEPYATNMI 437
 QY 421 GRYVNPFRKEMSSSTLNVLVAGAPRSQIVMLBASANTLLQDFCHAIKVGKXTQIIQ 480
 DB 438 GRYVNPFRKEMSSSTLNVLVAGAPRSQIVMLBASANTLLQDFCHAIKVGKXTQIIQ 497
 QY 481 SCVPISAVAGVAGLVTCTVFDLSGSKSDQVITAINIGIKDKNFMHYEPYATNMI 540
 DB 498 SCVPISAVAGVAGLVTCTVFDLSGSKSDQVITAINIGIKDKNFMHYEPYATNMI 557
 QY 541 QADIKLPGIPIKIWEAIOQASVAKKEILOIMNKTISKPRASKENGVPVETVOVPLSKR 600
 DB 558 QADIKLPGIPIKIWEAIOQASVAKKEILOIMNKTISKPRASKENGVPVETVOVPLSKR 617
 QY 601 AKFVPPGGVNLKQQAETGVITISQVDEETFSVPAPPSVMEHARDFITEICDDOEOOLE 660
 DB 618 AKFVPPGGVNLKQQAETGVITISQVDEETFSVPAPPSVMEHARDFITEICDDOEOOLE 677
 QY 661 FGAVTTATITEIRDGVNWKLYPNMTAVLHNTOLDNERL 700
 DB 678 FGAVTTATITEIRDGVNWKLYPNMTAVLHNTOLDNERL 717
 RESULT 3
 ABG30875 ID ABG30875 standard; Protein; 675 AA.
 AC ABG30875;
 DT 21-OCT-2002 (first entry)
 DE Human polynucleotide phosphorylase 74.25.
 EE Human polynucleotide phosphorylase 74.25; malignant tumour;
 KW haemopathy; human immunodeficiency virus infection; HIV;
 KW immunological disease; inflammation.
 OS Homo sapiens.
 PN CN1341720-A.
 PD 27-MAR-2002.
 PE 05-SEP-2000; 2000CN-0118992.
 PR 05-SEP-2000; 2000CN-0118992.

PA (SHAN-) SHANGHAI BIODOR GENE DEV CO LTD.
 XX Mao Y, Xie Y;
 PI WPI, 2002-501203/54.
 DR N-PSDB; ABR89228.
 XX New polynucleotide-phosphorylase 74.25 for treating
 PT malignant tumour, haemopathy, human immunodeficiency virus infection,
 PT immunological disease and various inflammations -
 XX Claim 1; Page 25-26 (Disclosure); 33pp; Chinese.
 PS The present invention discloses a new polypeptide-phosphorylase
 CC phosphorylase 74.25, a polynucleotide encoding the polypeptide and a
 CC method for producing the polypeptide using DNA recombination technology.
 CC The invention also discloses a method for curing several diseases, such
 CC as malignant tumour, haemopathy, human immunodeficiency virus (HIV)
 CC infection, immunological disease and various inflammations by using the
 CC polypeptide. The invention also discloses an antagonist for resisting the
 CC polypeptide and its therapeutic action, and also discloses the
 CC application of the polynucleotide encoding the new polynucleotide
 CC phosphorylase 74.25. The present sequence represents the
 CC polynucleotide phosphorylase 74.25.
 SO Sequence 675 AA:
 Query Match 85.2%; Score 3030; DB 23; Length 675;
 Best Local Similarity 94.5%; Pred. No. 5.8e-250;
 Matches 602; Conservative 3; Mismatches 4; Indels 28; Gaps 1;
 QY 64 MYTAVSKTKPSPOPMPLVVDYRQKAAAGRIPTNVLREPVSGDEILTSRIIDSIRP 123
 DB 1 MYTAVSKTKPSPOPMPLVVDYRQKAAAGRIPTNVLREPVSGDEILTSRIIDSIRP 60
 QY 124 LPPAGFYDTQVLCNLAVDGVNPDVLAINGASVALSLSDIPMNGPVAGVRIGIDSEY 183
 DB 61 LPPAGFYDTQVLCNLAVDGVNPDVLAINGASVALSLSDIPMNGPVAGVRIGIDSEY 120
 QY 184 VNPFRKEMSSSTLNVLVAGAPRSQIVMLBASANTLLQDFCHAIKVGKXTQIIQIG 243
 DB 121 VNPFRKEMSSSTLNVLVAGAPRSQIVMLBASANTLLQDFCHAIKVGKXTQIIQIG 180
 QY 244 QLVKRGVTKRTPQKLPSPSPVKTTHKLMERLAVPTDYHDKVSSDEAVNKIRLDT 303
 DB 181 QLVKRGVTKRTPQKLPSPSPVKTTHKLMERLAVPTDYHDKVSSDEAVNKIRLDT 240
 QY 304 EEOLEKEKPEADPEYELIESFNVAKEVRSIVLNEYKRCQGRDLTSLANVSCVDWFKTL 363
 DB 241 EEOLEKEKPEADPEYELIESFNVAKEVRSIVLNEYKRCQGRDLTSLANVSCVDWFKTL 300
 QY 364 HGSALFQRCQTOVLTCTVFDLSGSKSDQVITAINIGIKDKNFMHYEPYATNMI 423
 DB 301 HGSALFQRCQTOVLTCTVFDLSGSKSDQVITAINIGIKDKNFMHYEPYATNMI 360
 QY 424 TGLNRRREIGHALAEKALYPVLPDPPTITVTSSEVLESNSSSSMAASACGSLMDSGV 483
 DB 361 TGLNRRREIGHALAEKALYPVLPDPPTITVTSSEVLESNSSSSMAASACGSLMDSGV 420
 QY 484 PISAVAGVAGLVTCTVFDLSGSKSDQVITAINIGIKDKNFMHYEPYATNMI 543
 DB 421 PISAVAGVAGLVTCTVFDLSGSKSDQVITAINIGIKDKNFMHYEPYATNMI 480
 QY 544 IKLPGIPIKIWEAIOQASVAKKEILOIMNKTISKPRASKENGVPVETVOVPLSKRAKF 603
 DB 481 IKLPGIPIKIWEAIOQASVAKKEILOIMNKTISKPRASKENGVPVETVOVPLSKRAKF 527
 QY 604 VGPGGVNLKQQAETGVITISQVDEETFSVPAPPSVMEHARDFITEICDDOEOOLEGGA 663
 DB 528 VGPGGVNLKQQAETGVITISQVDEETFSVPAPPSVMEHARDFITEICDDOEOOLEGGA 572
 QY 664 VYATITEIRDGVNWKLYPNMTAVLHNTOLDNERL 700

DB 573 VYATITEIRDTGVWVKLYPMTAVLIANTOLDORXI 609

RESULT 4
AAB92684
ID AAB92684 standard; Protein; 504 AA.
XX
AC AAB92684;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:11065.
XX
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PE 28-JUN-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0189767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsubaki T;
XX
DR WPI; 2001-318749/34.
XX
PT primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11065; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 504 AA;
Query_Match 64.2%; Score 2285; DB 22; Length 504;
Best local similarity 100.0%; Pred. No. 1.9e-192;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGPFLPRRRALTLQVRLWSSAGSRAVAVDGNRKLKISSGKLARFADGSAVVOGSD 60

DB 18 DGPFLPRRRALTLQVRLWSSAGSRAVAVDGNRKLKISSGKLARFADGSAVVOGSD 77

QY 61 TAVVWTVASVKTKSPSPQFMPLVVDYROKAAAGRIPTNYLRREYGSDEKILRSRIIDRS 120
DB 78 TAVVWTVASVKTKSPSPQFMPLVVDYROKAAAGRIPTNYLRREYGSDEKILRSRIIDRS 137

QY 121 IRPLFPAGYFYDQVLCNLAVDGVNEPDVLAINGASVALSLSDIPNPGPVGAVRIGIID 180
DB 138 IRPLFPAGYFYDQVLCNLAVDGVNEPDVLAINGASVALSLSDIPNPGPVGAVRIGIID 197

QY 181 GEYVNPFRKEMSSSTNLNVAGAPKSIYMLASASENIIQDFCHAIKGVKXTQOIIQ 240
DB 198 GEYVNPFRKEMSSSTNLNVAGAPKSIYMLASASENIIQDFCHAIKGVKXTQOIIQ 257

QY 241 GIOOLVKEGTVTRTPQKLFPSPEIYKTHKLAEMELVAVPFDYEDHKVSRDEAVNKIR 300
DB 258 GIOOLVKEGTVTRTPQKLFPSPEIYKTHKLAEMELVAVPFDYEDHKVSRDEAVNKIR 317

QY 301 LDTEBQKEKFPADPYEIIIESFNVAKEVFRSIVLNEYRCGRDLTSLRNSCEVDMF 360
DB 318 LDTEBQKEKFPADPYEIIIESFNVAKEVFRSIVLNEYRCGRDLTSLRNSCEVDMF 377

QY 361 KTLHGSALFQRGQTQVLCVTFPDSLSGIRSDVITAINIGIKDKNFMALYEPYPATNEI 420
DB 378 KTLHGSALFQRGQTQVLCVTFPDSLSGIRSDVITAINIGIKDKNFMALYEPYPATNEI 437

QY 421 GKVTGNRRRELGHGALAEKALYPIVPRDFP 450
DB 438 GKVTGNRRRELGHGALAEKALYPIVPRDFP 467

RESULT 5
AAB08546
ID AAB08546 standard; Protein; 899 AA.
XX
AC AAB08546;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #8537.
XX
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN W0200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS72733.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 38905; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 899 AA;

Query Match 55.7%; Score 1982; DB 22; Length 899;
 Best Local Similarity 73.2%; Pred. No. 2.8e-165;
 Matches 408; Conservative 18; Mismatches 29; Indels 102; Gaps 2;

QY 10 DRALTOLOVRALMSSAGSRAVAVDIGNRKLKLEISSGKLARFADGSAVVOGDTAVMTAVS 69
 DB 440 DRALTOLOVRALMSSAGSRAVAVDIGNRKLKLEISSGKLARFADGSAVVOGDTAVMTAVS 499
 QY 70 KTKPSPOQFPLVVDYRQKAAAGRIPTNYLRREVGSDEKILTSRIIDRSIRPLFPAGY 129
 DB 500 KTKPSPOQFPLVVDYRQKAAAGRIPTNYLRREVGSDEKILTSRIIDRSIRPLFPAGY 521
 QY 130 FYDTQVLCNLAVDGVNBPDLAINGASVALSLSDIPMNGPVGAVRIGIIGDEYVNPTR 189
 DB 522 -----VGNTDDECVPNPR 535
 QY 190 KEMSSSTLNLVAVAGAPKSOIVMLEASAENILQODFCHAIKVGKVTQOIIQOIQOLVKET 249
 DB 536 KEMSSSTLNLVAVAGAPKSOIVMLEASAENILQODFCHAIKVGKVTQOIIQOIQOLVKET 595
 QY 250 GVTKTPQKLTTPSEIVKTYTHKLMERLYAVFTDYHDKVSRDAVVKIRLDTSEOLKE 309
 DB 596 GVTKTPQKLTTPSEIVKTYTHKLMERLYAVFTDYHDKVSRDAVVKIRLDTSEOLKE 655
 QY 310 KRPADPVEIIESEFVNAKEVRSIVLNEYKCDGDLTSLNVCCEVDMFTLHGSALF 369
 DB 656 IPEVDLVEIIESEFVNAKEVRSIVLNEYKCDGDLTSLNVCCEVDMFTLHGSALF 715
 QY 370 QRGOTQVLCVTFPDSLESGIKSDQYITAINGIKDNFMHYEPFYATNEIGKVTGLNR 429
 DB 716 QRGOTQVLCVTFPDSLESGIKSDQYITAINGIKDNFMHYEPFYATNEIGKVTGLNR 775
 QY 430 ELHGALAEKALPYIPRDPFTIRVTSVLESNSSSSMASACGSLALMDSGVPISSAV 489
 DB 776 ELHGALAEKALPYIPRDPFTIRVTSVLESNSSSSMASACGSLALMDSGVPISSAV 818
 QY 490 AGVALGLVTKTPDPERGEIEDYLLDIIIGIEPYNQDMPKLAGTKGTTALQADIKLPGI 549
 DB 819 -GVANGGLTKTPDPERGEIEDYLLDIIIGIEPYNQDMPKLAGTKGTTALQADIKLPGI 877
 QY 550 PIKIYMEAIQOASVAKK 566
 DB 878 TKIYMEAIQOASVAKK 894

RESULT 6
 ID ABG17275 standard; Protein; 899 AA.
 XX ABG17275;
 DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #17266.

XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN MO200175067-42.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HSE-) HXSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS81462.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 XX PT diagnostics, forensics, gene mapping, identification of mutations
 XX PT responsible for genetic disorders or other traits and to assess
 XX PT biodiversity

XX PS Claim 20; SEQ ID NO 47634; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
 XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 XX and gene mapping, and in recombinant production of (II). The
 XX polynucleotides are also used in diagnostics as expressed sequence tags
 XX for identifying expressed genes. (I) is useful in gene therapy techniques
 XX to restore normal activity of (II) or to treat disease states involving
 XX (II). (II) is useful for generating antibodies against it, detecting or
 XX quantitating a polypeptide in tissue, as molecular weight markers and as
 XX a food supplement. (II) and its binding partners are useful in medical
 XX imaging of sites expressing (II). (I) and (II) are useful for treating
 XX disorders involving aberrant protein expression or biological activity.
 XX The polypeptide and polynucleotide sequences have applications in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits to assess biodiversity
 XX and to produce other types of data and products dependent on DNA and
 XX amino acid sequences. ABG00010-ABG30377 represent novel human
 XX diagnostic amino acid sequences of the invention.
 XX Note: The sequence data for this patent did not appear in the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 899 AA;

Query Match 55.7%; Score 1982; DB 22; Length 899;
 Best Local Similarity 73.2%; Pred. No. 2.8e-165;
 Matches 408; Conservative 18; Mismatches 29; Indels 102; Gaps 2;

QY 10 DRALTOLOVRALMSSAGSRAVAVDIGNRKLKLEISSGKLARFADGSAVVOGDTAVMTAVS 69
 DB 440 DRALTOLOVRALMSSAGSRAVAVDIGNRKLKLEISSGKLARFADGSAVVOGDTAVMTAVS 499
 QY 70 KTKPSPOQFPLVVDYRQKAAAGRIPTNYLRREVGSDEKILTSRIIDRSIRPLFPAGY 129
 DB 500 KTKPSPOQFPLVVDYRQKAAAGRIPTNYLRREVGSDEKILTSRIIDRSIRPLFPAGY 521
 QY 130 FYDTQVLCNLAVDGVNBPDLAINGASVALSLSDIPMNGPVGAVRIGIIGDEYVNPTR 189
 DB 522 -----VGNTDDECVPNPR 535
 QY 190 KEMSSSTLNLVAVAGAPKSOIVMLEASAENILQODFCHAIKVGKVTQOIIQOIQOLVKET 249

Db 536 KEMSSSTLNWVAGAPKSGQVLEASASANILODFCHAIKVKYTOOIIQOLVKEI 595
 Qy 250 GVTKRTPOKLTFTSPSEIVKTYHKLAMERLVAVFTDYEDHKVSRDEAVNKIRLDTBEOLKE 309
 Db 596 GVTKRTPOKLTFTSPSEIVKTYHKLAMERLVAVFTDYEDHKVSRDEAVNKIRLDTBEOLKE 655
 Qy 310 KPEADPEYIIIESFNVAKVEFRSIVLNEYKRCGDRDITSLRNVSCEVDMFKTLHGSALF 369
 Db 656 IFPEVDLVEIIESFNVAKVEFRSIVLNEYKRCGDRDITSLRNVSCEVDMFKTLHGSALF 715
 Qy 370 ORGTQVLCYTPSPLESIGSDOYITAINGIKDKNFMFLHTEFPYATNEIGKTYGLNRR 429
 Db 716 ORGTQVLCYTPSPLESIGSDOYITAINGIKDKNFMFLHTEFPYATNEIGKTYGLNRR 775
 Qy 430 ELGHALAEKALVPIPRDPFTIRVTSEVLESNGSSMASAGCSGLALMDGVPISAV 489
 Db 776 ELGHALAEKALVPIPRDPFTIRVTSEVLESNGSSMASAGCSGLALMDGVPISAV 818
 Qy 490 AGVALGLVTKTDPKGEIEDYRLTLTGIEDYNGDMDFKLAGTNKGTALQADIKLPGI 549
 Db 819 -GVAMGLATKTDLKGEIEDYRLTLTGIEDYNGDMDFKLAGTNKGTALQADIKLPGI 877
 Qy 550 PIKIVMEAIQQAASVAKK 566
 Db 878 TMKIVMEAIQQAASVAKK 894

RESULT 7

ID ABG17276 standard; Protein; 439 AA.
 AC ABG17276;
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #17267.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 PN MO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PA (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS81463.
 FT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.
 PS Claim 20; SEQ ID NO 47635; 103bp; English.
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG0377 represent novel human diagnostic amino acid sequences of the invention.
 Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 439 AA;

Query Match 53.9%; Score 1917; DB 22; Length 439;
 Best Local Similarity 78.7%; Pred. No. 4.6e-160;
 Matches 396; Conservative 6; Mismatches 15; Indels 86; Gaps 3;

Qy 1 DGPFLLPRDRALTOLOVRALMSSAGSRAVAVDLGNRKLEISSGKLARPADGSAVVOGSD 60
 Db 18 DGPFLLPRDRALTOLOVRALMSSAGSRAVAVDLGNRKLEISSGKLARPADGSAVVOGSD 77
 Qy 61 TAVVTVAVSKTKSPSQFMPVVDYRQKAAAGRIPTNYLRREVTSDEKILTSRIIDRS 120
 Db 78 TTVVTVAVSKTKSPSQFMPVVDYRQKAAAGRIPTNYLRREVTSDEKILTSRIIDRS 108
 Qy 121 IRPLFPAGYFYDQVLCNLAVDGVNEPDYLAINGASVALSLSDIPWNGCVGAVRIGIID 180
 Db 109 -----AVRIGIID 116

Qy 181 GEYVNPTRKEMSSSTLNWVAGAPKSGQVLEASASANILODFCHAIKVKYTOOIIQ 240
 Db 117 GEYVNPTRKEMSSSTLNWVAGAPKSGQVLEASASANILODFCHAIKVKYTOOIIQ 176
 Qy 241 GIOQLVKEIGVTKTPKLTFTSPSEIVKTYHKLAMERLVAVFTDYEDHKVSRDEAVNKIR 300
 Db 177 GIOQLVKEIGVTKTPKLTFTSPSEIVKTYHKLAMERLVAVFTDYEDHKVSRDEAVNKIR 236

Qy 301 LDTEBQLEKPEADPEYIIIESFNVAKVEFRSIVLNEYKRCGDRDITSLRNVSCEVDMF 360
 Db 237 LDTEBQLEKPEADPEYIIIESFNVAKVEFRSIVLNEYKRCGDRDITSLRNVSCEVDMF 296
 Qy 361 KTLHGSALFORGTQVLCYTPSPLESIGSDOYITAINGIKDKNFMFLHTEFPYATNEI 420
 Db 297 KTLHGSALFORGTQVLCYTPSPLESIGSDOYITAINGIKDKNFMFLHTEFPYATNEI 356

Qy 421 GKVTGLNRRELGHGALAEKALVPIPRDPFTIRVTSEVLESNGSSMASAGCSGLALMD 480
 Db 357 GKVTGLNRRELGHGALAEKALVPIPRDPFTIRVTSEVLESNGSSMASAGCSGLALMD 414
 Qy 481 SGVPISAVA---GVALGLVTKT 500
 Db 415 NGFRGSNFIICCRPYKGLVTKT 437

RESULT 8

ID ABB58546 standard; Protein; 748 AA.
 AC ABB58546;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 2430.
 KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
 OS Drosophila melanogaster.
 PN MO200171042-A2.

XX 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2001; 2000US-191637P.
 PF 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 XX (PEK) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EM;
 DR WPI; 2001-656860/75.
 XX N-PSDB; ABL02649.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Disclosure; SEQ ID NO 2430; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB116175), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (AB101840-AB116175).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 748 AA:

Query Match 51.8%; Score 1842.5; DB 22; Length 748;
 Best Local Similarity 52.8%; Pred. No. 4.1e-153;
 Matches 364; Conservative 124; Mismatches 170; Indels 31; Gaps 6;

19 RALMSAGSRAVAVDL--GRRKRLSSGKARFADGSAVVGSGTAVVAVTAVSKTRSP 75
 29 RGIQSSSEAPSVENFNGRMTSSGRLRFPAGTAVCMGTAVVAVTAVAKKPNP 88
 76 SQ-FMPLVVDYRQKAAAGRIPTNYLRREVSTDEKILTSRIIDRSIRLPAGYFYDQ 134
 89 CGGFPPLVVDYRQKAAAGRIPTNYLRREVSTDEKILTSRIIDRSIRLPAGYFYDQ 148
 135 VLCNLLAVDGNVDPVLAINGASVALSLSDIPMNGPVGAIRIGIIDEVYVNPTRKMS 194
 149 LVCNMLAMDVAHSPDVLAINAASWALSLSIDIPMNGPVGAVRGVLCDEGVLINPTREIQ 208
 195 STLANVAVGAPKSOIVMLEASANIIOQPCFAIKVGVKTYQIIQGIQQLVKEGVTYK 254
 209 SOLDLIVASTKONLVLMGKGNVLIQDILKAIKGTREAFIIRHRLKAKVGRQR 268
 255 TPQKLFPSPELVKTKHKLAMERLVAVFTDYHDKVSRDEAVNKIRLDTBEOLKCKPRA 314
 269 EREVAALVDEPELGAVKRSCERLKEIFDSTHDKSRNANAVASNIYDVMSFPPT 328
 315 DEYRIIESFNVAKEVFRSIVANEKRCGDRDLSIRNVSCVDEWFKTLHGSALFORQGT 374
 329 EBSLITEGNGTSRTIFRELIFERGLRCGRDYDQLRNISCVDMWKPLHGSALFORQGT 388
 375 QVLCTVTPSLSEGIKSDOVITAIN--GIKDKNPMHLYFPYATNIEKVTGLARRREG 432
 389 QVCFVTSIDSSQSSAMKLDLS--LALASGGLKAKNFMHLYFPYATNIEKVTGLARRREG 447
 433 HGALAEKALPYPIPRDFPTIRVTSVLESGSSSSMASACGSLALMBGVPISAVAGV 492
 448 HGALEKRLSLPTLPNDYFTVTLTSVLESGSSSSMASACGSLALMBGVPISAVAGV 507
 493 AIGLVTK-TDPEKGEIEDYRLITDILGIEDYNGDMDFKIAGTNKGIITLQADIKLPGIPI 551

DB 508 AIGLVTKPEENDTKHLDYRLITDILGIEDYNGDM-----M 544
 QY KIWEAIOASVAKKEILQINMKTSKPRASRKGNGPVETQVPLSKRAKVGPGYNL 611
 DB 545 KVMESLQKADPAIDNSNIDMSSEALREPRKPKSMPSEITLVPPQPAULIGSGLM 604
 QY 612 KKLQAEVTGVTISQVDEETFSVAPPTSVHMRAPFITEICDDQEQLEFGAVYATITE 671
 DB 605 KRIYLETSTSLAVDETHFNVPAPSOAMDRAKELISGVYKERVDPLEFGIYTAKITE 664
 QY 672 IRDTGVAVKLYPNMTAVILHNTQDNERL 700
 DB 665 LRDTEGVAVILYSPMPALHNSOLDQRKI 693
 RESULT 9
 ABG08547
 ID ABG08547 standard; Protein; 541 AA.
 XX
 XX ABG08547;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #8538.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PF WO200175067-A2.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HSE)- HXSEQ INC.
 XX
 PI Dmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 XX
 DR N-PSDB; AAS72734.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID NO 38906; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO


```

Db 398 RAGRRREIGHGALGERALKXIIIPDTADPPTIRIVSEVESNGSSSQASICGSTALLMDA 457
Qy 482 GVPISAVAGVAGIAGVTKTDPKEGEIEDYRLTDIIGIEDYNGMDPFKIAGTKGITALQ 541
Db 458 GVPKAPVAGIAGMGLVTRD-----STYLTIDIGMEDALGMDPFKAGTKEGITALQ 510
Qy 542 ADIKLPGIPIKIWEALIQASVAKKEILOIMNKTISKPRASKENGPPVETVOVPLSKRA 601
Db 511 MDIKIDGLTREIIEBALQARRGRLEIMNMLQTIQPTRELSAVAPKVTMTIKDKIR 570
Qy 602 KTVGPGGYNLKLQATGVTISQVDEETSVFAPPTSVMHEARDPTTEICDDQOQLEF 661
Db 571 DVIIGPGKKINIEIDETGVKLDIEOGTIFIGAVDAQMINRAREIIEITR-----EAEV 625
Qy 662 GAVYATITEIRDGTGMVKLYPNMTAVLHNTQDNERL 700
Db 626 GQTYQATVKRIEKYGAFLPGKDA-LHHSIQISKRI 663

RESULT 11
AU36732
ID AU36732 standard; Protein; 698 AA.
XX
AC AU36732;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #902.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KM antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN W0200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WC-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
XX
PR 23-MAY-2000; 2000US-206848P.
XX
PR 26-MAY-2000; 2000US-207727P.
XX
PR 23-OCT-2000; 2000US-242578P.
XX
PR 27-NOV-2000; 2000US-253625P.
XX
PR 22-DEC-2000; 2000US-257931P.
XX
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haasebeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto R, Xu H;
XX
DR WPI; 2001-611495/70.
XX
DR N-PSDB; AAS54591.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 12325; 511p; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery

```

```

CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 698 AA.
Query Match 33.5%; Score 1191.5; DB 22; Length 698;
Best Local Similarity 39.2%; Pred. No. 1e-95;
Matches 274; Conservative 117; Mismatches 249; Indels 59; Gaps 12;

Qy 24 SAGSAAVAVDLNKRKLEISGCLAPADGSAAVVGSDTAVMTAVSKTPSSQFMPLV 83
Db 2 SECKVVFTEWAGRSILTEGTQLAKQANGAVRVYDVTVASTAVASKPRDGDFFPLTV 61
Qy 84 DYKRAAAAGRIPTYVLRREVGTSDKEILTSRIIDRSIRPLFPAGYFVDTVLNLAVD 143
Db 62 NYEEKVYAGKLPGGFKRREGPRDATTILRPLRPLPKGKHQVQIMMWLSAD 121
Qy 144 GYNEBDVLAINGASVALSLSDIPNNGPVGAVRIGIIDEYVNPTRKEMSSSTLNLVVG 203
Db 122 PDCSPQAMAMIGSSMALSVSDIPFGPIAGVAVGIDGKIINPTEKEVSRDLDEVAG 181
Qy 204 APKSOIWLKSAENILQDFCHAIKV-----VKTQOIIQGIQQLVKETGYTKTP 256
Db 182 -HKDAVNVWVEGASITEQEMLEAIFGHEIRQLRVLDVFOQIVDHIQPKOE----- 232
Qy 257 QKLFPS-----PEIVKYTHKLAMERLVANFTYEHDKYSRDAVANKIRLDTEQKE 309
Db 223 ---FIPARDEALVERIKSLTEKGLKETVLT-----DKOORDENLDLKK--BEIYVE 281
Qy 310 KPEPADP-----YEIISFNVVAKVEPRSVLVNMYKRCDDGLTSLNVSCEVDMFKTL 363
Db 282 FIDEEDPENELLKEVVALINELVKEVRRLADKIRPDGKKPBEIRPLDSEVGILPRT 341
Qy 364 HGSALFQGGQVQLCTVTPFDSIESGKSDQVYTAINGIKDKNFMLYHEPPYATNEIGV 423
Db 342 HSSGLFTRGQTOALSVLTLGAL---GDYQILDIGLGPREEKRFMHVNFPMPSVGETSPV 397
Qy 424 TGLNRREIGHALAEKALYVIP--RDPEPTIRVTSYLTESNGSSSMASAGCSIALMDS 481
Db 398 RAGRRREIGHGALGERALKXIIIPDTADPPTIRIVSEVESNGSSSQASICGSTALLMDA 457
Qy 482 GVPISAVAGVAGIAGVTKTDPKEGEIEDYRLTDIIGIEDYNGMDPFKIAGTKGITALQ 541
Db 458 GVPKAPVAGIAGMGLVTRD-----STYLTIDIGMEDALGMDPFKAGTKEGITALQ 510
Qy 542 ADIKLPGIPIKIWEALIQASVAKKEILOIMNKTISKPRASKENGPPVETVOVPLSKRA 601
Db 511 MDIKIDGLTREIIEBALQARRGRLEIMNMLQTIQPTRELSAVAPKVTMTIKDKIR 570
Qy 602 KTVGPGGYNLKLQATGVTISQVDEETSVFAPPTSVMHEARDPTTEICDDQOQLEF 661
Db 571 DVIIGPGKKINIEIDETGVKLDIEOGTIFIGAVDAQMINRAREIIEITR-----EAEV 625
Qy 662 GAVYATITEIRDGTGMVKLYPNMTAVLHNTQDNERL 700
Db 626 GQTYQATVKRIEKYGAFLPGKDA-LHHSIQISKRI 663

RESULT 12
ABP40063
ID ABP40063 standard; Protein; 706 AA.
XX
AC ABP40063;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4908.

```

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KM antibacterial; gene therapy.
 XX
 OS Staphylococcus epidermidis.
 XX
 FN US6380370-B1.
 XX
 PD 30-APR-2002.
 XX
 PF 13-AUG-1998; 98US-0134001.
 XX
 PR 14-AUG-1997; 97US-055779P.
 PR 08-NOV-1997; 97US-064964P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 FI Doucette-Stamm LA, Bush D;
 DR WPI; 2002-381255/41.
 XX N-PSDB; ABN92608.
 XX
 PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 XX
 PS Disclosure; SEQ ID 4908; 267pp; English.
 XX
 CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP3124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life
 CC cycle or inhibit S. epidermidis infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.
 CC
 SQ Sequence 706 AA;
 Query Match 32.8%; Score 1165.5; DB 23; Length 706;
 Best Local Similarity 38.7%; Pred. No. 2.1e-93;
 Matches 270; Conservative 123; Mismatches 250; Indels 55; Gaps 13;
 QY 24 SAGRAVAVDGNKLTSSGKLARFADGSAVVGSGRAVAVTAVSKTSPSQFMPVYV 83
 DB 7 SQEKVFKTEWAGSLTLETQQLAKQAGAVLVKRGDVTVAATASKKEPRDGFPLTV 66
 QY 84 DYKOKAAAAGRIPTVYLRREVGTSDKEILTSRIIDRSIRLPFGYFYDTQVLCNLAVD 143
 DB 67 NYEKMTAAAGKIPGFKRREGRPDEALTLRLIDRIPRLFPKGYRHDVGINNIVLSAD 126
 QY 144 GVNEPDVLAINGASVALSLDIPMNGPVGAIRIGIIDEYVNPTRKEMSSSTLNLVAG 203
 DB 127 PDCSPPEMAAMIGSSMALSVSDIPFGQPIAGVNVGIDDKYVINSVADKEISRDLLEVAG 186
 QY 204 APRSQIWLKASANIILQODFCHAIKVG-----VKTQQLIGIQQLVETGVTKRP 256
 DB 187 -HKAVNVNVEAGSELTSEMLELIFGHEIKRLVAQOQIIDIHQIKOE----- 237
 QY 257 QKLFTP---SPEIVYTHKLAMER--LYAVFTDYEHDKVSDAEVANKIRLTTEQLKEKF 311
 DB 238 ---FVPEREDDLVKKVSLTFEDKGLKOTVLT---FDKQGRDNDMDALK---EEVVGHFL 288
 QY 312 PEADP-----YEIESFNVAKEVFRSIVLNEYKRCQGRDLTSLRNVSCVDMFKTLHG 365
 DB 289 DEEDPENETLVKEVYAILNDLIKESVRLIADKIRPDGRKVRDEIRPLESEVGLLPRAHG 348
 QY 366 SALTQKQGVTLCTVTPSLSEGIKSDGVTAINGIKDKNMMLHYEPFPVATNIEIKVYG 425
 DB 349 SGLFRTQOTLQSVLTGLAL---GDYQLIDGLDPEVKKRPMHHYFNPFSVGETGPVRA 404

QY 426 LNRRELGHALAEKALYVIP--RDPEFTIRVTSEVLESGSSSMASACGSLALMDSGV 483
 DB 405 PGRREIGHALGERALLIIPDQDFPYTIRIVSEVLESGSSSQASICGSLTALMDAGV 464
 QY 484 PISSAVGVAIGVTKTDPEKGEIEDYRLTLIDIGIDYNGDMDFKAGNKGTALQAD 543
 DB 465 PIKAPVAGIAMGLVTRD-----SYTILTDIQGMDALGDMDFKAGTDDGITAIQMD 517
 QY 544 IKLPGIPIKIWEALIQASVAKKEILQIMNKTSIKPPASRKENGVPVETVQVPLSKAKF 603
 DB 518 IKIDGLREVIEVEALDQROGRALIMDMHTIEQPRBELSAYAPKVTNINSNDKIRDV 577
 QY 604 VPGGVNMLKQLQAEVTGISQVDETFESVAPTPSVNHEARDFITEICKDQEQOLEFGA 663
 DB 578 IGPQKKINEIDETGVKLDIEQGTIFIGAVDQAMINRAKEITEDTR-----EAEVQG 632
 QY 664 VYTTATTEIRIDTGVNKLTPMTAVLALHNTQLDNERLN 701
 DB 633 VYHAKVRIEKYGAFVELFPGKDA-LIHISQISOERIN 669

RESULT 13

AAU38175 ID AU38175 standard; Protein; 721 AA.

AAU38175; AC

DT 14-FEB-2002 (first entry)

XX Salmonella typhi cellular proliferation protein #66.

KM Antisense; prokaryotic cellular proliferation protein;
 KM antibiotic; antibacterial; drug design.

OS Salmonella typhi.

FN WO200170955-A2.

PD 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2001; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207727P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253625P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PI Haselebeck R, Ohlsen KL, Zyekind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 DR WPI; 2001-611495/70.
 DR N-PSDB; AAS56034.PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 13768; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.

The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.
 Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_jct_sequences.

Sequence 721 AA;

Query Match 32.6%; Score 1158; DB 22; Length 721;
 Best Local Similarity 39.9%; Pred. No. 9, 8e-93;
 Matches 270; Conservative 117; Mismatches 250; Indels 40; Gaps 11;

35 GNRKLEISSGLARPADGSAVVOGDTAVMTAVSTKRSPO-FMPLVVDYRQAAAG 93
 22 GCHTYTLETGMARQATAVWVSMDDTAVFTVVGKAKKPODFPPLVNYOERTYAG 81
 94 RPTVYLRREVGTSDKEILTSRIIDRSIRPLFPAGFYDVTQNLAAVDGNBPVLA 153
 82 RIFGSPKRRGRBSEETILARLIDRPVPLPFGFVAVVIAVVSVPVNPVIVAM 141
 154 NGASVALSLSDIPMNGPVGAVRIGIIDERYVNPTRKENSSTLNLVAGAPKQIVME 213
 142 IGASVALSLSGIPFGPICAARVGYINDQVLPQDELKSKLDLVAGT-BAVLAWE 200
 214 ASAENILQODFCHAIKVGKTKQIIGIQQLVKGTGVTKTPQKLPFSPPIVK----- 268
 201 SEAEILSEPTMLGAVVFGHEQOOVIAINDLVKEXGKPRMWO-----PEAVDALNA 254
 269 YTHKLAMERL---YAVFTDYEHDKVSRDEAVNKRILDTREOLKEKPEADPEIIESENV 325
 255 RVALAASLSLDAIKI---TDKQERYAQOVVYKSETTEQLAIBETLIDANELSHLHA 309
 336 VAKVEFERSIVINEYKRCGRDLTSLRNVSCEVDMFKTLHGSALFORGOTOVLCTVTPSL 385
 310 IEKNVVRSLVLAGPRIDREKDMIRGLDVRGTGLPRTHGSLLPFRGETQALVTNTLGT- 368
 386 ESGISDOVTTAINGIKDKNFKMLHTEPPPYATNEIGKVTGLNRLRELGHGALAEKALPYI 445
 369---ARDQVLDLMEGERTSFLFHNFPYSGVETGMVSPKRRREIGHRLAKRGVLAVM 425
 446 P--RDEPTIRYTSVLESGSSMAACGSLALMDSVPTSSAVAGALGLVTKTPE 503
 426 PDMKFPYTVRVVSEITBSNGSSMAVCGASLALMDAGVPIKAAGVAGLAMEVREGD-- 483
 504 KGEIEDYRLTDLIGIEDYNGDMDFKIAGTNKGTALQADILPGIPIKIWEAIQASV 563
 484-----NYVVLSDILDEBHLGMDKRVAGSRDGLALQMDIKIBGITEIMQVVALNOKG 538
 564 AKGEILQIMNKTISKPRASRKENGVPVETVOVPLSKRAKFGVPGGVNLKQLQAEVTGIS 623
 539 ARHLIGVWEQAINAPRGDISFAPIRIHITIKISTKIKOVIGKGSVIRALTEETGITE 598
 624 QVDEFTSFVAPFSPVMEHARDPITEICDDDEQOLERGAVVTATITTEBDTGWVAKXP 683
 599 IEDDTIVKIAATDEKAKYAIRIBI-----TAIEVGRIYNSKVTIRVDGAFVAIGG 653
 684 NMTAVLHNTQDNERL 700
 654 GKEG-LVHISQIADKRV 669

RESULT 14
 ID AA03792 standard; Protein; 1034 AA.

AA03792;
 AA03792;
 11-JUN-1999 (first entry)
 DT
 XX

DB S. aureus polypeptide.
 XX Staphylococcus aureus polypeptide; thyroiditis; infective carditis;
 XX lung abscess; secretory diarrhoea; cerebral abscess; conjunctivitis;
 XX toxic shock syndrome; folliculitis; septic arthritis; antibacterial;
 XX H. pylori infection; gastric ulcer; adenocarcinoma.
 OS Staphylococcus aureus.
 XX EP905243-A2.
 XX 31-MAR-1999.
 XX 03-AUG-1998; 98EP-0306185.
 XX 05-AUG-1997; 97US-005387.
 XX (SMIK) SMITHLINE BEECHAM CORP.
 XX (SMIK) SMITHLINE BEECHAM PLC.
 XX Burnham MKR, Lonetto MA, Warren PV;
 XX WPI, 1999-192667/17.
 XX N-PSDB; AAX31862.
 XX New essential polypeptides from Staphylococcus aureus useful for
 XX treating diseases such as infective endocarditis and toxic shock
 XX syndrome
 XX Claim 31; Page 50-55; 70pp; English.
 XX
 XX The invention provides new Staphylococcus aureus polypeptides
 XX (AA03781-94) and the gene (AAX31851-864) encoding them. Host cells
 XX containing vectors comprising the nucleic acid sequences are used for the
 XX recombinant expression of the proteins. The polypeptides can be used to
 XX screen for modulators for use in antibacterial therapy. The polypeptides,
 XX their antagonists and agonists are used to prevent or treat diseases
 XX caused by S. aureus such as thyroiditis, lung abscesses, infective
 XX carditis, secretory diarrhoea, cerebral abscesses, conjunctivitis, toxic
 XX shock syndrome folliculitis and septic arthritis. Screening for the
 XX presence of the polypeptides may be used to diagnose, predict the
 XX susceptibility to, or stage the progress of these S. aureus diseases and
 XX diseases caused by Helicobacter pylori such as gastric ulcers and gastric
 XX adenocarcinoma. There is not much information known about the essential
 XX genes expressed by S. aureus during infection but these new polypeptides
 XX have been identified as essential. They can therefore be used to develop
 XX antibacterial compounds specific for those essential genes and this
 XX ensures the effectiveness of the compounds in killing S. aureus. In
 XX addition, these polypeptides can be used to effectively diagnose and
 XX treat infections and diseases caused by S. aureus without the risk of
 XX development of antibiotic resistance. The present sequence represents a
 XX S. aureus polypeptide which has homology to a polyribonucleotide
 XX nucleotidyltransferase.
 XX
 XX Sequence 1034 AA;
 XX
 XX Query Match 32.5%; Score 1156; DB 20; Length 1034;
 XX Best Local Similarity 39.3%; Pred. No. 2, 7e-92;
 XX Matches 265; Conservative 113; Mismatches 239; Indels 58; Gaps 11;

24 SAGRAVAVDVGNRKLEISSGLARPADGSAVVOGDTAVMTAVSTKRSPOFMPLV 83
 2 SOEKVFTFWAGSLTETQOLAKONGAVLVGDTVUSTATASKEPDDGDFPLTV 61
 84 DYRQAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGFYDVTQNLAAVD 143
 62 NYEERMYAAGRIPOGFKRGRBPGDDATLTARLIDRPVPLPFGYKXDVQIMNVVLSAD 121
 144 GVNBPVLAINGASVALSLSDIPMNGPVGAVRIGIIDERYVNPTRKENSSTLNLVAG 203
 122 PDGSPQAMAMGSSMALSVSDIPQGPVAGVNGVIGDKYINLPVEKEVSRDLVLEAG 181
 204 APKSQIWLASASENILQODFCHAIKVG-----VKYTIQIIGIQQLVKGTGVTKTP 256

Db 182 -HKQAVNVEAGASEITEOEMLAIFGHEEIQRLVDFQOQIIVDHIOGVKOE----- 232
Qy 257 QKFTPS-----PEIVKTHKLAMERLAVFVDYEHDKYSRREAVNKRRLDTEBOJKE 309
Db 233 ---PAPARDEALVERIKSLTEEGKLETTLT-----DKOORDENLNLK---BEIVNE 281
Qy 310 KFPBAPD-----YEIIESFNVAKEVRSIVLNEYKRCDDRDLTSLANVSCVDMFKTL 363
Db 282 FIDEDPENELLIEVAALINELVKEEVRRLIADBKIRPDQRKDEIRLPDSEVGIILPRT 341
Qy 364 HGSALFQROGTOVCTTVPDSLESGIKSDQVITAINGIKQNFMLHFEPPYATNEIGKV 423
Db 342 HGSGLFTFGQTOALSVTLGAL-----GDYQILDIGGPEEERFPMHNVFPMFSGETGPV 397
Qy 424 TGLMRRELGHGALAEKLLVPIP---RDPFTIRVTSKLTESNGSSSMASACGSLALMDS 481
Db 398 RAPGRREIGHALGRALKYIIPRTADPPYIRIVSEVLSNGSSSSQMSIGSTLALMDA 457
Qy 482 GVPISSAVAGVAGLVTDEPEKEIEDYRLITDILGIEDYNGDMDFKIACTNGKITPAQ 541
Db 458 GVPIKAPVAGIAGLVTRD-----SYTILTDIOGMDALGMDDFVAGTKEGISTAIQ 510
Qy 542 ADIKLPGIPIKIWEALIQOASVANKELIQINNKTIKSPKRSKENGVPVETVQVPLSKRA 601
Db 511 MDIKIDGLREIIEEALQARGRLIEMNHMLQITIDQRTIELSAVAPKVTMTIKPKIR 570
Qy 602 KFPVPGGYNLKKLAETGVITISQVDEEFVSFAPRPSVMEHARDFITICDDDOEQULEF 661
Db 571 DVIPGPKKINEIIDEETGVKLDISODGTIFGAVDQAMINARBIIEITR-----EAEV 625
Qy 662 GAVVTATITETIRDTG 676
Db 626 GQTYQATYKRIEKYG 640

RESULT 15
AAU34719
ID AAU34719 standard; Protein; 734 AA.

AC AAU34719;

DT 14-FEB-2002 (first entry)

DE E. coli cellular proliferation protein #300.

KM Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

OS Escherichia coli.

XX WO200170955-A2.

XX 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawlock JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR N-PSDB; AAS52578.

PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX Example 3; Seq ID No 10312; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 734 AA;

Query Match 32.5%; Score 1155; DB 22; Length 734;

Best Local Similarity 39.9%; Pred. No. 1.9e-92;

Matches 269; Conservative 110; Mismatches 252; Indels 36; Gaps 11;

Qy 35 GNRKLEISSGKLARFADSAVAVOSGDTAVWAVTAVSKRPPSQ-FMPLVVDYRQKAAAG 93

Db 35 GQHTVLETGMARQATAAVAVMSMDTAVFVTVGQKARPGDFPLTVNYQERTAAAG 94

Qy 94 RIPTNYLRREVGSDEKILTSRIIDRSIRPLFPAGYFDTQVLCNLAVDGNEDVLA 153

Db 95 RIPGSPFRREGRPSEGETTLARLIDRPIRLPFGFENFENQVAVTAVSVNPQVPDIVAM 154

Qy 154 NGASVALSLDIPNNGVAVRIGIIDGEEVYVNTREKSSSTNLNVVAGAPKQIYMLE 213

Db 155 IGASALSLSGIFPNPGPIGAARVGYINDQVLANPTQDLBSKLDLVVAGT-EAAVLNVE 213

Qy 214 ASAEIILQODFCHAIKGVYVYQOIIQGIQOLVETGVTKTPQKLTFTSP---EIVKYT 270

Db 214 SEAQLSBEDQMLGAVVGHGQOVVIONINELVNEAG---KPRWDQPEVNEALNARV 269

Qy 271 HKLAMERL---YAVFTDYEHDKVSRDEAVNKRILDTBOJKEKFPBAPYHIIISFNVA 327

Db 270 AALAEARLSDAVRI-----TDKQERYAQQVDVYIKSEFTATILAEDETLIDENELGELIAIE 324

Qy 328 KEVPRSTVINEYKRCDDRDLTSLANVSCVDMFKTLGSLAFQROGTOVCTVPDSLES 387

Db 325 KNVVRSTVNLAGEPRIDRBERKDMIRGLDVRGTGLPRTGSALEFTRGETQALVTALGT-- 381

Qy 388 GIKSDQVITAINGIKQNFMLHFEPPYATNEIGYTGALNRRELGHGALAEKALVPIP- 446

Db 382 -ARDAQVLDLMEGRITPTLFHYNFPPYSVGETGVMSPKRRELGHRLAKRGVLAAMPD 440

Qy 447 -RDPFTIRVTSVLEBSNGSSSMASACGSLALMDSGVPISSAVAGVAGLVTTPDEKG 505

Db 441 MDRPFTYRVVSVITTESNGSSSMASVCGASIALMDAGVPIKAAVAGIAGLVKEGD--- 496

Qy 506 EIDYRLITDILGIEDYNGDMDFKIACTNGKITLQADIKLPGIPIKIWEALIQOASVAK 565

Db 497 ---NYVLASDILGDEHLDGMDDFVAGSRDGIISALQMDIKIEGTHKEIMOVALQAGAR 553

Qy 566 KEIIOINNKTIKSPKRSKENGVPVETVQVPLSKRAKFGVGNLKKLQAEVTITISQV 625

Db 554 LHILGWEGALINARSGIISFAPRIHITIKINPDKIXOYIGKGSVIALTETITIEIE 613

Qy 626 DEETFSVAPPTPSVMEHARDFITEICDDDOEQULEFGAVVTATITETIRDTGVMVKLYPNM 685

Db 614 DDGTAKIAATDEKAKHAIRIEEI-----TAEIEVGRVYTGKTRIVDFGAFVAIGGK 668
Oy 686 TAVLHNTQJDNERL 700
|:|:|:|:|:|:
Db 669 EG-LVHISQIADKRV 682

Search completed: January 8, 2004, 10:38:47
Job time : 51 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2004, 10:37:57 ; Search time 22 Seconds
(without alignments)
1355.871 Million cell updates/sec

Title: US-09-907-907a-42

Perfect score: 3557
Sequence: 1 DDPFLPRDRALTOQVRA.....TAVLHNTQDNERLITLLP 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfillseq.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1165.5	32.8	706	4 US-09-134-001C-4908	Sequence 4908, Ap
2	1107	31.1	750	4 US-09-107-532A-5868	Sequence 5868, Ap
3	1065	29.9	705	4 US-09-328-352-8112	Sequence 8112, Ap
4	976.5	27.5	697	4 US-09-252-991A-27283	Sequence 27283, A
5	884	24.9	568	4 US-09-198-452A-1073	Sequence 1073, Ap
6	226	6.4	127	4 US-09-198-452A-1074	Sequence 1074, Ap
7	140.5	3.9	1416	4 US-09-071-035-404	Sequence 404, Ap
8	140.5	3.9	252	4 US-09-252-991A-2776	Sequence 402, App
9	140	3.9	191	4 US-09-252-991A-27186	Sequence 27186, A
10	137	3.9	244	4 US-09-328-352-5628	Sequence 5628, Ap
11	128.5	3.6	871	4 US-09-134-001C-3579	Sequence 3579, Ap
12	127.5	3.6	471	4 US-09-107-532A-5705	Sequence 5705, Ap
13	124.5	3.5	1306	3 US-08-999-774A-13	Sequence 13, Appl
14	124.5	3.5	1780	1 US-08-769-309A-5	Sequence 5, Appl
15	116	3.3	875	4 US-08-094-570-5	Sequence 4537, Ap
16	116	3.3	875	4 US-09-107-532A-4537	Sequence 594, App
17	115	3.2	648	4 US-09-198-452A-584	Sequence 28, Appl
18	114	3.2	2270	4 US-09-581-909-3	Sequence 5687, Ap
19	114	3.2	235	4 US-08-747-562-28	Sequence 2, Appl
20	113	3.2	745	4 US-09-107-532A-5887	Sequence 2, Appl
21	112.5	3.1	519	3 US-08-997-445D-2	Sequence 2, Appl
22	112	3.1	644	1 US-08-021-608D-2	Sequence 2, Appl
23	109	3.1	644	1 US-08-726-160-2	Sequence 2, Appl
24	109	3.1	644	1 PCT-US94-01782-2	Sequence 2, Appl
25	109	3.1	644	1 US-08-476-008-42	Sequence 42, Appl
26	108.5	3.1	428	1 US-08-306-063-42	Sequence 42, Appl
27	108.5	3.1	428	1 US-08-306-063-42	Sequence 42, Appl

28	108.5	3.1	428	1 US-08-833-485-42	Sequence 42, Appl
29	108.5	3.1	428	4 US-09-137-440-42	Sequence 42, Appl
30	108.5	3.1	2504	4 US-09-328-352-5821	Sequence 5821, Ap
31	108	3.0	693	1 US-08-463-620-11	Sequence 11, Appl
32	108	3.0	693	1 US-08-224-917-11	Sequence 11, Appl
33	108	3.0	693	2 US-08-914-853-11	Sequence 11, Appl
34	108	3.0	693	5 PCT-US95-03934A-11	Sequence 11, Appl
35	108	3.0	959	4 US-09-914-259-67	Sequence 67, Appl
36	107.5	3.0	961	4 US-09-914-259-66	Sequence 66, Appl
37	107	3.0	836	1 US-08-426-627-6	Sequence 6, Appl
38	106.5	3.0	837	1 US-08-426-627-23	Sequence 23, Appl
39	106.5	3.0	1118	4 US-09-379-523-3	Sequence 3, Appl
40	106	3.0	579	4 US-09-643-593-348	Sequence 348, App
41	106	3.0	579	4 US-09-542-615A-348	Sequence 348, App
42	106	3.0	579	4 US-09-606-421B-348	Sequence 348, App
43	105	3.0	444	4 US-09-107-532A-3661	Sequence 3661, App
44	105	3.0	533	3 US-09-603-185-2	Sequence 2, Appl
45	105	3.0	937	1 US-08-253-155A-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-09-134-001C-4908
; Sequence 4908, Application US/09134001C
; Patent No. 6180370
; GENERAL INFORMATION:
; APPLICANT: Lynx Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4908
; LENGTH: 706
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4908

Query Match	32.8%	Score 1165.5;	DB 4;	Length 706;
Best Local Similarity	38.7%	Pred. No. 2e-100;		
Matches 270;	Conservative 123;	Mismatches 250;	Indels 55;	Gaps 13;
QY	24	SAGSAVAVDGNRLTSSGTLAPFADGSAVVOGDPFANWTVAVSKTKRPSQFMPVIV	83	
DB	7	SOEKVFTFWAGRLTLETQLAQKNGAVLVIRGDTVAVSTAVASKEPRDGFPLTV	66	
QY	84	DYRQAAAGRIPTVYLRREVTSDEKILTSRIIDRSIRPLFPAGFYDTQVLCMLAVD	143	
DB	67	NYEHCYAAAGTIPGFKRGRGPRGDEATLRIIRPLFPKRYRDVQIMNIVLSAD	126	
QY	144	GVNEEDVIALNGASVALSLSDIPMNGPYGAVIRIGIIDEVYVNPFRKMSSTLNLVAVG	203	
DB	127	PDCSEMAAMIGSVALSVSDIPFGPIAGVAVGIDKYYVNPVADKEISRDLLEVAG	186	
QY	204	APKSIIVLLEASAEHIILODFCHAIKVG-----VKYTOQIQIOQLVETGYTKRP	256	
DB	187	HKDVNVNVEAGASITSEMEALFPQHEETKRLVAQOEIIDIHIOPIKE-----	237	
QY	257	QKLFTPT---SPEIVYTKHLMER--LYAVFTDEYHDVSRDEAVNKRRLDTEQKEKF	311	
DB	238	---FPPVERDDIVKYSVLTEDKGLKDTVLT---FDQQRDENIDALK---EAVVGHFL	288	
QY	312	PEADP-----YEIIESFNVAKEVFRSIVTNEYKRCGRDLTSLRNVSCVDMFKTLHG	365	
DB	289	DEDPENETLVKRYVAIINDLIKESVRLIADKIRPDGRVDEIRPLESEVGLPRAHG	348	

366 SALFQRCQTVLCTVTPDLSLGKSDVITAINGIKDKNFMHYEPVATNEIGKVTG 425
349 SGLFTRGQVLAIVLTIGAL-----GDYQLDGLGPEVEKRMHNFNFVSGETGPRA 404
426 LNRELHGLAALKALYVIP--RDPFTITVTSVLESGSSMAACGSLALMDSGV 483
405 PGRREIGHGALGERLRYIIPDTPFTIRIVSVLESGSSSQASICGSLTLMADGV 464
484 PISAVAGVAGLVTKTPEKGEIEDVRLTDIIGIEDYNGMDPKIAGTKGITALOAD 543
465 PIKAPVAGIAGLVTTRD-----STILTIDIGEMALGDMDFKAGITDITAILMD 517
544 IKLPGPIKIVMEAIQASVAKELIQMKTISKPRASRENGPVETVQVPLSKAKF 603
518 IKIDOLFREVIEALQARQRLAIMDHMLHTIEQPREELSAVAPKVATMSINPKIRDV 577
604 VGPBGYNLKLQAEIGVTISQVDETFEVPAPTSVMHEADPTTEICKDOEQOLEGA 663
578 IGPBGKINELIIDEFGVLDIEODGTIFIGAVDQAMINRAKEIIEDITR-----EAEVQ 632
664 VYTAITEIRDTGVVVKLYPMNTAVILHNTOLDNERLN 701
633 VYHAKVKRIEKYGAFLPGKDA-LVHISQISQERIN 669

RESULT 2

US-09-107-532A-5868
Sequence 5868, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-8277
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5868:
SEQUENCE CHARACTERISTICS:
LENGTH: 750 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...750

SEQUENCE DESCRIPTION: SEQ ID NO: 5868:
US-09-107-532A-5868

Query Match 31.1%; Score 1107; DB 4; Length 750;
Best Local Similarity 36.8%; Pred. No. 7e-95;
Matches 250; Conservative 144; Mismatches 247; Indels 38; Gaps 12;

35 GNRRLRISGKLAFADSAVQSGDPAVAVTVSKRPSPGPMPLVVDYRQKAAAGR 94
54 GGRFLQLEVOQLKQKNGAVLVKGVDTIVLSAIVASLEADOTDFPLITIVSEGMVAK 113
95 IPTNYLRREVSTSDKEILTSRIIDRSIRPLPPAGYFYDTOLCNLAADVNEBDVLAIN 154
114 IPGFIRRESRPSTFALTLRLIDRPIRMFAGFRRVQVTINVMVSKTDCSPAMAAML 173
155 GASVALSLDIPNMGPGAVRIGIIOSEYVNPFRKMSSTLNTLVVAGPKSQIVWLEA 214
174 GSSLALSISDIPDGPAGVGVGRVNGSVVLPVFEQAEQTDLELTVAAGT-KQALNIVES 232
215 SAENILQODFCHAIKVGVKTYTOOLIQGLVKEGVTKETPOKLFPSPEIVYTHKLA 274
233 GAKVSEEDMLGALPFDALKEIVAFQEBIVQAVG--KERMVTLQVDEVLK--KEI 287
275 MERLY-----AVFTDYEDHVRSDPAVVKIRLDEQLKEKP--EADPY---ELIESFN 324
288 FDASYATMKAAMVTE--EKLAREDNIEQKIDIREAVAKRFAGHEDHDLKEVKQITE 344
325 VVAKVRSIVLVNEYKACDGRDLTSLRVNVCEDMFTLGSALPORGQOVLTCTVPDS 384
345 DLEKDVVRELITIDKIRPDKRKLDEIRPLSEVSLPRVHSGLFTTGQVLSACTL-- 402
385 LESGKSDVITAINGIKDKNFMHYEPVATNEIGKTVGLNRRELHGLAALKALYPV 444
403 --APLGRHQLIDGLGVSVSKRFTHNYFPQSVGSTRASPGRRREGHGLGERALAQV 460
445 IP--RDPFTIRYTSVLESGSSMAACGSLALMDSVPISSAVAGVAGLVTTDP 502
461 IPSEBEPYTRIVAAEVLBSGSSQASICAGTLALMDAGVPIKAPVAGIAMGLVS---- 516
503 EKEIEDYRLTDLIGIEDYNGMDPKIAGTNNGITLQMDILPGIPKIVMEAIQOAS 562
517 --DGEVYTLTIDIGLEHDLGMDPFVAGTQGITLQMDIKIQTGQITLQALQAK 573
563 VAKKEILOINMKTISKPRASRENGPVETVQVPLSKRAKFGVGNLKKLOAETGVTI 622
574 QARWEILBELSTIAPRELSQYAPKIEMLQLEPAIKIVIGGSGTNGIIDEIVKI 633
623 SQVDETFEVPAPTSVMHEADPTTEICKDOEQOLEFGAVYATITTEIRDGVVVKLY 682
634 DIDDGKVSISADAEWIKKAIKIEDLTK-----EVKGVSVYLGKVVRIEFGAFVNL 688
683 PNMTAVILHNTOLDNERLN 701
689 KGKDG-LVHISQISQERIN 706

RESULT 3

US-09-328-352-8112
Sequence 8112, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 8112
LENGTH: 705
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-8112


```

Sequence 1073, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
TITLE OF INVENTION: and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 1073
LENGTH: 568
TYPE: PR1
ORGANISM: Chlamydia pneumoniae
FEATURE:
NAME/KEY: SITE
LOCATION: 1...568
OTHER INFORMATION: Xaa=unknown or other

```

US-09-198-452A-1073

Query Match 24.9%; Score 884; DB 4; Length 568;

Best Local Similarity 36.9%; Pred. No. 3.9e-74; Indels 40; Gaps 13;

Matches 205; Conservative 112; Mismatches 198; Indels 40; Gaps 13;

QY 157 SVALSLDIPMNGPVGAVRIGIIDEYVNPTRKEMSSSTLNLVAGAPKQIWLLEASA 216
 DB 8 SALLAISDIPQSNIVAGRICIDINQWVINTKTELASTLDLVLAGT-ENALIMEBHC 66
 QY 217 ENLQDPCCHAIKGVKVTQOIIQIQLVKEGTGKTPQKLPSPSEIYKTHKAME 276
 DB 67 DFTPEQVLDIAIEFHKHAIVTICKQLQMOEVEGSKMLSAVPLPA-EVLTAVKECQD 125
 QY 277 RLYAVFTDYEDHKVSRDAVNNKIRL-DTEBOLKEKFPADPEIIESFNVA-----KE 329
 DB 126 KFTLEFN-----IKDKVHAATHHEENILELQREDD-DLPSSNIIQACKTILKSD 177
 QY 330 VFRSIVLNEYKRCGRDLTSLRNVSCVDMFKTLHGSALFORQOTVL-CTVTPDSLES 387
 DB 178 TMRALIRDRERIRADGRSLTTRPITIFTSYLPRTGSCILFRTGETQLAVCTLGSEAMA- 236
 QY 388 GIKSQVITAINGIQDKFPMHYEPFPATNEIGKVTGLNRELGHGLAEKALYVLP- 446
 DB 237 -----QRYEDINGEGLSKFYIQLYFFPPPSVGEVGIQSGRREIGHGLAEKALSHALPD 291
 QY 447 -RDPFTIRVTSEVLESNGSSMASACGSLALMDSGVPISSAVAGVAGIYTKTPDEKG 505
 DB 292 SATPPTIRISNTIESNGSSMASVCGCLALMDAGVPISSPIGIMGLL--DDQG 348
 QY 506 EIBDYRLITLIGLEDYNGDMFKIAGTNKGTALQADIKLPGIPIKIMEAIQASVAK 565
 DB 349 AI-----ILSDISGLDHDIGDMFKIAGSGKITTAQMDIKVEGIPALMKKALSQAKGC 404
 QY 566 KEILOINAKTISKPRASKENGPPVETQVPLSKRAKVVGGVNLKQLQMETGTISQV 625
 DB 405 NDILINMEALSAPEKADLSQYAPRIETWQIKPTKIASYIGGGKQIRQIIBETGVQIDVN 464
 QY 626 DEFTSVAPTPSVNHEARDFTIEICKDOEQLEFGAVTATITEIDTGVNVLQYPM 685
 DB 465 DLGVVISASASAINKAKEILEGLV-----GEVAVGTTIKRVTISVAFGAFVGLDPK 519
 QY 686 TAVLHNTQLDNERL 700
 DB 520 EG-LCHISECSRQRI 533

RESULT 6
 US-09-198-452A-1074
 ; Sequence 1074, Application US/09198452A
 ; Patent No. 6559294
 ; GENERAL INFORMATION:
 ; APPLICANT: Griffiths, R.
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
 ; FILE REFERENCE: 9710-003-999
 ; CURRENT APPLICATION NUMBER: US/09/198,452A
 ; CURRENT FILING DATE: 1998-11-24
 ; NUMBER OF SEQ ID NOS: 6849
 ; SEQ ID NO 1074
 ; LENGTH: 127
 ; TYPE: PRT
 ; ORGANISM: Chlamydia pneumoniae
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: 1...127
 ; OTHER INFORMATION: Xaa=unknown or other
 US-09-198-452A-1074

Query Match 6.4%; Score 226; DB 4; Length 127;
 Best Local Similarity 42.6%; Pred. No. 1.9e-13;
 Matches 49; Conservative 23; Mismatches 43; Indels 0; Gaps 0;

QY 39 LEISSGLARFADSGAVVSGDFTAVMTAVSKTSPSPQMPFLVVDYRQKAAAGRIPTN 98
 DB 7 LVFTGTIAIQANAVAVRSRGTCVFPASACAVDDDKVDLPLRVDYQEKFSSTGKTLG 66
 QY 99 YLRREVTSDEKILTSRIIDRSIRPLFPAGYFYDTQVLCNLAVGVNBPDLAI 153
 DB 67 FIKREGPSEKEILVSRILDRSLRSPSPYRLMQVQVLSYVWSYDQVLPDLAI 121

RESULT 7

US-09-071-035-404
 ; Sequence 404, Application US/09071035
 ; Patent No. 6448043
 ; GENERAL INFORMATION:
 ; APPLICANT: Gil H. Choi
 ; TITLE OF INVENTION: Enterococcus faecalis polynucleotides and polypeptides
 ; NUMBER OF SEQUENCES: 496
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/071,035
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: A. Anders Brookes
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB369P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 404:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1416 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-071-035-404

Query Match 3.9%; Score 140.5; DB 4; Length 1416;
 Best Local Similarity 21.1%; Pred. No. 0.0014;
 Matches 135; Conservative 79; Mismatches 236; Indels 189; Gaps 31;

QY 158 VALSLDIPMNGPVGAVRIGIIDEY-----VNPTRKEMSSSTLNLVVA 202
 DB 481 INLSLST-TPVIGPEKAIQ-LVSDQYIEPISVNPPLAETAWGNYDONGAVSSRTTAVSM 537
 QY 203 GAPKQIWLLEASA--NILO----ODECHAIKGVKVTQOIIQIQLVKEGTGKTP 256
 DB 538 GSKKEPIQNLLEIKYKHNNYLSLRATKTIYFYKLGDTYTV-----TP 579
 QY 257 QKLTPEPEIYKTHKLAERLYAVFTDYEDHKVSRDAVNNKIRLDTBEOLEKFPADP 316
 DB 580 ----TSQGSVYKFTTPTTNEIQIPIGFNVPDLSPKDXX--IPVDI----- 619
 QY 317 YEIESFNVAKEVFRSIVLNEYKRCGRDLTSLRN--VSCVDMFKTLHGSALFORQ 373
 DB 620 --IPITWSAGLTPVDVTVTNTSKSGSERTLOSKNQFLVVARNDSPDSLSVRTKIPAG- 676

QY 374 TQVLTCTVFDLSGSIKSDQVITAINGIKDKNFMHYPFPATNEIGKVTGLNRELGH 433
 Db 677 ---ADVLFDIYD--VSNDQVDSIYPQYWDG--QYFDKPTMTNSPG----- 715
 QY 434 GALAEKALYPIVPRD-----PFP-----TIRVTEVLESNGS 465
 Db 716 ---YPTITPENTNSYTFDFGKTKRKYIIEYKANGMIDVPTLYITGTAKEPQSN 767
 QY 466 SSMSACGSLALMDSGVPISSA--VAGVAIGLVTKDPEKGEIRD--YRLITDILGIED 521
 Db 768 NNEGSA---SVSVQNEBALDILSATQANPTLKNKVTITVTKNIDKTKHRYKNPTIEL-- 822
 QY 522 YNGDMDFKIAGTNGKITLALD---IKLPGIP--IKIYWEAIQASVAKKEILQIMNKTI 576
 Db 823 ---TPKGTINAQIDLSITVGVPEDAVSLEKTKNGAKVIFKQVTLTENITI 871
 QY 577 SKPRASRENGPVVEV-----QVPLSKRAKFGVPGGYNLKQLOAETGVITIS----- 623
 Db 872 EYNTVSANAGQIYETTTIDSETLNMASAKKVTAPITLKPEGBAGIIVIAATATPYT 931
 QY 624 -QVDEETFSV---FAPTPSVMEHARDPITEICDQDQOQLEFGAV---YTATITEIRD 674
 Db 932 HNVEDENQAIKVSFELIDNVHTATETFT---DEKQYSFDALMTGDTLRLVTVNQ 986
 QY 675 ---TGVMVKLY--PMTAVLLHNTQDNERLNI 702
 Db 987 EYSVDEEYLTKAIKLVKGDNLKIPLTXT-IDHSRLQV 1024

RESULT 8

US-09-071-035-402
 ; Sequence 402, Application US/09071035
 ; Patent No. 6448043

GENERAL INFORMATION:
 APPLICANT: Gil H. Choi
 TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 NUMBER OF SEQUENCES: 496
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44M storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/071,035
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: A. Anders Brookes
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB369P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 402:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1448 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-071-035-402

Query Match 3.9%; Score 140.5; DB 4; Length 1448;

Best Local Similarity 21.1%; Pred. No. 0.0014;
 Matches 135; Conservative 79; Mismatches 236; Indels 189; Gaps 31;
 QY 158 VALSIDIPNPGVCAVAGIIGIDSEY-----VANPTREMS-----SSTLNLVYA 202
 Db 508 INLSLS-TPVIGPNALQ--LVSDQYIPISVPLNEMETGANGDQAVSSRTVSWM 564
 QY 203 GAPKQIWMLEASAE--NILQ---QDFCHAIKGVKVTQGIQGIQQLVETGVTKPT 256
 Db 555 GSKKEPIQNLKVKHPYLSIRATKEIYFYKLGIDYTV-----TP 606
 QY 257 OKLFTPSBEIYKYTKLMEBLVAVFTDYEHDKVSDEAVNKIRLDTEOLKEKPEADP 316
 Db 607 ---TSDGSVAKFTTPIINELIPIGFVYVPSLPSKDS--IPVDT----- 646
 QY 317 YEIISFNVVAKVEVRSIYLVBYKRCQDRLTSLN--VSCVUMFTIHGSLFQSQ 373
 Db 647 --IPITMSAEGLPVDITVTNNSKRSERLTQSSNQPLVNARNSFSLSVRTKIPAG- 703
 QY 374 TQVLTCTVFDLSGSIKSDQVITAINGIKDKNFMHYPFPATNEIGKVTGLNRELGH 433
 Db 704 ---ADVLFDIYD--VSNDQVDSIYPQYWDG--QYFDKPTMTNSPG----- 742
 QY 434 GALAEKALYPIVPRD-----PFP-----TIRVTEVLESNGS 465
 Db 743 ---YPTITPENTNSYTFDFGKTKRKYIIEYKANGMIDVPTLYITGTAKEPQSN 794
 QY 466 SSMSACGSLALMDSGVPISSA--VAGVAIGLVTKDPEKGEIRD--YRLITDILGIED 521
 Db 795 NNEGSA---SVSVQNEBALDILSATQANPTLKNKVTITVTKNIDKTKHRYKNPTIEL-- 849
 QY 522 YNGDMDFKIAGTNGKITLALD---IKLPGIP--IKIYWEAIQASVAKKEILQIMNKTI 576
 Db 850 ---TPKGTINAQIDLSITVGVPEDAVSLEKTKNGAKVIFKQVTLTENITI 898
 QY 577 SKPRASRENGPVVEV-----QVPLSKRAKFGVPGGYNLKQLOAETGVITIS----- 623
 Db 899 EYNTVSANAGQIYETTTIDSETLNMASAKKVTAPITLKPEGBAGIIVIAATATPYT 958
 QY 624 -QVDEETFSV---FAPTPSVMEHARDPITEICDQDQOQLEFGAV---YTATITEIRD 674
 Db 959 HNVEDENQAIKVSFELIDNVHTATETFT---DEKQYSFDALMTGDTLRLVTVNQ 1013
 QY 675 ---TGVMVKLY--PMTAVLLHNTQDNERLNI 702
 Db 1014 EYSVDEEYLTKAIKLVKGDNLKIPLTXT-IDHSRLQV 1051

RESULT 9

US-09-252-991A-23776
 ; Sequence 23776, Application US/09252991A
 ; Patent No. 6551795

GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 PRIOR FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 23776
 LENGTH: 252
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-23776

Query Match 3.9%; Score 140; DB 4; Length 252;
 Best Local Similarity 22.1%; Pred. No. 7.7e-05;
 Matches 62; Conservative 48; Mismatches 109; Indels 62; Gaps 11;

QY 271 HGLAMERLYAVTDEHDKVSRDEAV-NKIRLDTREQLKEKPEADPYIIIESFNVVAKK 329
 DB 496 LEKAAELQGTIPQEKELQEFEEAFQDGTGDESEMIRVVSDEIGIVGWT----- 550
 QY 330 VRSJVLNEYKCKDGDRLTSLRNVSCEVDMFTLHGSALFQNGQTOVLCCTVPDSLESCI 389
 DB 551 ---GIPVSLVTEREKLSLSDI-----LHNRVV---GQDKAVDLV----- 586
 QY 390 KSDOYITAINGIKDKNFMHYEPFPAATMEIGV-----TGLNRRRLGALAEKALYVP 444
 DB 587 -SDAVVRAAGIKDPN-----RPISSFLFGPTGVGKELA-----KSLASS 627
 QY 445 IPRDEPFTIRV-TSEVLNESNGSSSMASACGSLALMDSVPISSAVAGVAGLVTKTDE 503
 DB 628 LFDSEKMRIRIMSEYMEKHAVSRIGAPGYG-HDEGGQJLEAVRRNPYSVILLDEVE 686
 QY 504 KEIEDRYLLTDLG---IEDYNG-DMDK---IAGTKGITALQADIKLPG---IP 551
 DB 687 KASHDFVNLQILDEGRLLTDSKGRSVDFKNTIIMTSNIGSQVLEENYKXAGEISDDTE 746
 QY 552 KIVMEAIQOASVAKKEILLIMNKTISKPRASRKNENGVETVQVPLSKR----- 600
 DB 747 KAVMDSLR-AVFKPEILNRMDIYLFKPLSVDDMSMIVDKITLQNMRLDQHSIEVT 804
 QY 601 ---AKFVGPQYV-----LKKL---QAEQV-----TISQVD-----EETP 630
 DB 805 BEAKWLGSEAVEPQRCARPLKRFVQRIETPIARMIKESLPBGIIIVDLNDKELDF 864
 QY 631 SVPAFT 636
 DB 865 KVKKEPT 870

RESULT 13

US-09-107-532A-5705
 ; Sequence 5705, Application US/09107532A
 ; Patent No. 6583275
 ; GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02354
 COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: PC
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09107,532A
 FILING DATE: 30-Jun-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/085,598
 FILING DATE: 14 May 1998
 APPLICATION NUMBER: 60/051571
 FILING DATE: July 2, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Atinello, Pamela Deneke
 REGISTRATION NUMBER: 40,489
 REFERENCE/DOCKET NUMBER: GTC-012
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781)893-5007
 TELEFAX: (781)893-8277
 INFORMATION FOR SEQ ID NO: 5705:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 471 amino acids
 TYPE: amino acid

TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (B) LOCATION 1...471
 SEQUENCE DESCRIPTION: SEQ ID NO: 5705:
 US-09-107-532A-5705

Query Match 3.5%; Score 124.5; DB 4; Length 471;
 Best Local Similarity 21.1%; Pred. No. 0.0065;
 Matches 66; Conservative 56; Mismatches 106; Indels 85; Gaps 15;

QY 37 RKLISISGLARFADGSAVNGSDTAVMNTAVSKTRPSPQMPLVYDRQAAAAARLP 96
 DB 32 RLDEIQT-NVFKYPPGSAVIRGDTTVICSAIENSVP--FL-----RETGTGWT 80
 QY 97 TNY--LRREVGTSDKE-----ILTSRIIDRSIRPLPAGYFYDTQVL--CNLLA 141
 DB 81 AYSMLPRTATSTRNRBSKGLFGRTMEIORLGRSIRAVVDLEKLGERSIIVDCVYQ 140
 QY 142 VGVNEPDLAINGASVAL-----SLSDIPNNGPVGAVRIGII-DGEYVNPTR 189
 DB 141 ADGGR--TASITGAFVALKLAIEKLEKESEDPRIEHLAAVSGVILPDGTCVTDLDY 198
 QY 190 KEMSSSTNLVYAGAPKQIYMLEASANILOQDFCAIKYGVKTTQIIQ--IQLV 246
 DB 199 QEDSALYDVMILVMTESKGFVEIQGTGE-----BATFPDEQINEMLFPKNAIEDLI 250
 QY 247 KETGVTKRTPOKLFTSPPEIYVYTHKLAMERLYAVFTDYEHDKVSDEAVNIRLDTS-- 304
 DB 251 KE-----QK-----HALTFQAND-ERISFTKIITATNP 281
 QY 305 ---EQLEKPEPA 314
 DB 282 GKAEERFMFKEA 294

RESULT 14

US-08-999-774A-13
 ; Sequence 13, Application US/08999774A
 ; Patent No. 6274312
 ; GENERAL INFORMATION:
 APPLICANT: Gish, Kurt C.
 APPLICANT: Seghezzi, Wolfgang
 APPLICANT: Shanahan, Frances
 APPLICANT: Lees, Emma M.
 APPLICANT: McClanahan, Terrill K.
 TITLE OF INVENTION: Intercellular Regulatory Molecules;
 TITLE OF INVENTION: Related Reagents
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: DNAX Research Institute
 STREET: 901 California Avenue
 CITY: Palo Alto
 STATE: California
 COUNTRY: USA
 ZIP: 94304-1104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/999,774A
 FILING DATE: 10-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/032,818
 FILING DATE: 11-DEC-1996
 ATTORNEY/AGENT INFORMATION:

US-08-999-774A-13
 ; Sequence 13, Application US/08999774A
 ; Patent No. 6274312
 ; GENERAL INFORMATION:
 APPLICANT: Gish, Kurt C.
 APPLICANT: Seghezzi, Wolfgang
 APPLICANT: Shanahan, Frances
 APPLICANT: Lees, Emma M.
 APPLICANT: McClanahan, Terrill K.
 TITLE OF INVENTION: Intercellular Regulatory Molecules;
 TITLE OF INVENTION: Related Reagents
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: DNAX Research Institute
 STREET: 901 California Avenue
 CITY: Palo Alto
 STATE: California
 COUNTRY: USA
 ZIP: 94304-1104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/999,774A
 FILING DATE: 10-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/032,818
 FILING DATE: 11-DEC-1996
 ATTORNEY/AGENT INFORMATION:

NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0646
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)852-9196
TELEFAX: (650)496-1200
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1306 amino acids
TYPE: amino acid
STRANDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-999-774A-13

Query Match 3.5%; Score 124.5; DB 3; Length 1306;
Best Local Similarity 19.7%; Pred. No. 0.038; Indels 257; Gaps 38;
Matches 149; Conservative 112; Mismatches 239;
QY 73 PPSQFMELVVDYRQKAAAGRIPTNYRREVGTSDEK-ILTSRIIDR-----SIRPL 124
DB 530 PLFNQILFLMS-----PI-----LEDQERHLLVVKYIDILVYKLDLVRP- 569
QY 125 PPAGFYDTQVLCNLAVDGVNEPVLAINGASVALSLSDIPWNGPVGAVRIGIIDG--E 182
DB 570 ----YVHKILVIVIEPLID--EDYVARVEGLEIISNLAKAAGLATMTWSTWRPDIDNDE 622
QY 183 YVNPPTREMSSTLNLVVA---GAPKQIWMLEASANIILQODPCHAIKVKVKTQO-- 237
DB 623 YVRITTAAPA-----VVASNLGIP-SLPLPLKAVCKS--KSWQARHTGIKIVQOIA 672
QY 238 ----IIQGIQQLVKETGVTKRTPQKLFPSPEIVKYTHKLMERLYAVFTDYEHKVS 291
DB 673 ILMGAILPHLRSLV-----ELI-----EHGLVD 696
QY 292 RDEAVNKRILTEBOLKEKPEADPYELIESNNVAK----- 328
DB 697 EQQKRTISALAIAMAE---AATPYG-IESFDSVLKPLMKGIROHRSGLAFLKAIQY 752
QY 329 -----EVPRSIVLMEYKRCGRDLTSLRNVSCEV-- 357
DB 753 LIPLMDAEVANYTREVMILIRFQSPDEEMKVLVLYVQCCGTGVEANYITETILP 812
QY 358 DMFKTL--HGSALFORQTOVLCTVFPDSLESIGIKSDOVITAI-NGIKDKNFMHYERP 414
DB 813 PFFHGFQWHRMALLDRNRYRQLVDTTVELASNKVGAELISRIYDLDLDE---AEQYRK 868
QY 415 YATMEIGVTVG-LNRRELGHALAEKALYPIVPRDPTTIRVTSVLESNGSSMASA-- 471
DB 869 MVMETIEKIMGNLGAADIDH-KLEBQLIDGIL--YAFQOETTESVLMNGFGTVVNALG 924
QY 472 -----CGGSLALMDGVPISSAVAGVALGVTTDPEKGBIEYRLLTDL-LEIE 520
DB 925 KRVRKYLPOICGTVMRLNNK--SAKYRQQAADIDSHRAVVMKTQCEBKMLGHGLVLY 981
QY 521 DYNQDMDFKIAGTNKGTALQADILPGI-----PIKIVM-----EAIQO--- 560
DB 982 EYLGEYEVGLSIG--ALKALIVNVIGMHKTPPIKOLLRLPLILKNRHEKVGENDID 1039
QY 561 -----ASVAKKELIQTINKTKISKRPRSKKE-NGPVVETVQVPLSKAKFYGPQGY 609
DB 1040 LVGRADRGAEVVSAREMMRICFELLELLKAKHKIRATVNT---FGYAKAIGPBDV 1095
QY 610 -----NLKQAOE---TGVTISQVDETPSVFAPSPVMEHAR----- 644
DB 1096 LATLLNNKLVQDRQNVCTVAIAIV-AETCSPFTVLPLALMNEYVPELNVQNGVLSLS 1154
QY 645 ---DFTTEICKDDQOLEFGAVYATITTEIRDTGVM 678
DB 1155 FLFEYIGEMGKD-----YIVAVTPLEDLALM 1180

RESULT 15

US-08-769-309A-5
Sequence 5, Application US/08769309A
Patent No. 5741890
GENERAL INFORMATION:
APPLICANT: Scott, John D.,
APPLICANT: Nauer, Brian J.,
APPLICANT: Klauck, Theresa M.,
TITLE OF INVENTION: Protein Binding Domains of Gravin
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower/233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/769,309A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5741890and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-0448
TELEFAX: 312-474-6300
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-769-309A-5

Query Match 3.3%; Score 116; DB 1; Length 1780;
Best Local Similarity 19.6%; Pred. No. 0.41;
Matches 134; Conservative 99; Mismatches 269; Indels 182; Gaps 35;
QY 40 EISSGKLARFADGAVVQSGDTAVMTAVASKTSPQFMELVVDYRQKAAAGRIPTNY 99
DB 882 ELSESQVHMMMA--AAVADGTRATTI---EERSPSWISASVTE-----PLEQ 923
QY 100 LRREVGTSDEKELITSRIIDRSIRPLPAGYFYDTQVLCNLAVDG---VNEPVL--AIN 154
DB 924 VEAAALLTREVLREVIABERPTV-----TEPLPENREARGDTVVSABLTPRAVT 976
QY 155 GASVALSLSDIPWNGPVGAVRIGIIDSEYVNPFRKEMSSSTLNLVYAGAPKSOIWMLEA 214
DB 977 AAETA-----GPIGS-----EEGTASAABETTMV-----SAVSQTLTD 1010
QY 215 SAENILQODPCHAIKVQV-----KYTQOIIQGIQQLVKETGVTKXT--POKLFPSPE 265
DB 1011 SPDTTEERTPVQVEGGVGDPIEEOERTQEVLAQVAKVEESOLPGTGEBEDVLP--- 1067
QY 266 IVKTHKLMERLYAVFTDYEHDKVSRDEAVNKRILDTBOLKEKPEADPYELI----- 320
DB 1068 VQRAEARPREQQAASGKKEITDV-----LKVDAQEKATEPFTQG--KVVGQTP 1116
QY 321 ESFNVVAKYEFRSIVLNEYKRCGRDLTSLRNVSCEVDMFKTLHGSALFORQTOVLCTV 380
DB 1117 ESFE-KAPQVYESISESELV-----TTQQA-----VLSQEMV----- 1152
QY 381 TFDLESIGISDQVITAINIGIKDKNFMHYEPFYATNEIGKVGLNRELGH-----GA 435
DB 1153 ---MEQALPPDSVETPTDSTGTPV-ADFDAPGTQKDEIVIEHEENVHLVPRGT 1207

```

Qy 436 LAE-----KALYVIRPDPPFTIRVTSEVLESNGSSMASACGSLALMDGVPISAVAG 491
Db 1208 EAEAVPAQKERPPAPSSFFVFC-----EETKOSKMD-----TLEHTDKEVSVE-----1252
Qy 492 VAIGLVTKTDPEKEIBDY--RLTDLIGIEBYNGDMDEKINAGTKGITALQADIKLPGI 549
Db 1253 --VSTLSKTEGTQ-EADQYADEKTKVDPPEGLBSIDTGITVSRKVT---EVALKG-1304
Qy 550 PIKIYMEALQASVAKKEILQIMNKTISKPRASRKENGVPVETVQVPLSKRAKAVGPGGY 609
Db 1305 -----EGTEBAECKKDDALEQSHAKSP-----SPVEREMVQVREKTEAEPHY 1351
Qy 610 NUKKLOAETGVTSQ-----VD--EETFSY-FAPTPSVMEHARDFTTEICK 652
Db 1352 NEEKLEHETAVTVSESVSKOLLQTVNVPIIDGAKEVSLBSPPCLGQE-----EAVCT 1406
Qy 653 DDQEQQLFERGAYTATITEIRDTG 676
Db 1407 KIQVOSSEASFTLTAAAEKVLG 1430

```

Search completed: January 8, 2004, 10:41:27
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2004, 10:38:52 ; Search time 37 Seconds
(without alignments)
3839.578 Million cell updates/sec

Title: US-09-907-907A-42

Perfect score: 3557
Sequence: 1 DGFLLPRDRALTOLOVRA.....TAVLHNTQDNERILNLLP 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCT05_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3557	100.0	705	11	US-09-907-907A-42
2	3557	100.0	705	11	US-09-907-907A-44
3	1191.5	33.5	696	9	US-09-815-242-5443
4	1191.5	33.5	698	9	US-09-815-242-12325
5	1158	32.6	721	9	US-09-815-242-13768
6	1155	32.5	734	9	US-09-815-242-10312
7	1134	31.9	709	9	US-09-815-242-10989
8	1132	31.8	705	11	US-09-907-907A-43
9	1128.5	31.7	704	9	US-09-815-242-10882
10	1127	31.7	702	9	US-09-815-242-4975
11	1097	30.8	613	9	US-09-815-242-11755
12	1097	30.8	701	9	US-09-815-242-12050
13	1054	29.6	737	9	US-09-815-242-13699
14	1014.5	28.5	753	10	US-09-738-626-5665
15	999.5	28.1	738	15	US-10-156-761-10060

16	997.5	28.0	752	10	US-09-712-363-252	Sequence 252, App
17	831.5	23.4	688	9	US-09-815-242-11416	Sequence 11416, A
18	831.5	23.4	688	12	US-09-882-227-284	Sequence 284, App
19	829.5	23.3	688	9	US-09-815-242-11576	Sequence 11576, A
20	708	19.9	164	10	US-09-925-300-11451	Sequence 11451, A
21	543	15.3	188	10	US-09-764-864-1513	Sequence 1115, App
22	287	8.1	85	10	US-09-764-864-1513	Sequence 1533, App
23	175	4.9	245	11	US-09-907-907A-45	Sequence 45, App
24	147	4.1	451	9	US-09-815-242-10721	Sequence 10721, A
25	143	4.0	245	10	US-09-738-626-6245	Sequence 6245, App
26	139	3.9	921	9	US-09-800-729-199	Sequence 199, App
27	136.5	3.8	245	11	US-09-893-519A-48	Sequence 48, App
28	136	3.8	240	12	US-10-032-585-7234	Sequence 7234, App
29	136	3.8	244	15	US-10-156-761-12704	Sequence 12704, A
30	131	3.7	256	15	US-10-102-806-664	Sequence 664, App
31	131	3.7	964	15	US-10-156-761-9998	Sequence 9998, App
32	129.5	3.6	863	12	US-10-369-493-10297	Sequence 10297, A
33	126	3.5	691	9	US-09-815-242-4977	Sequence 4977, App
34	126	3.5	693	9	US-09-815-242-10481	Sequence 10481, A
35	126	3.5	816	12	US-10-369-493-18348	Sequence 18348, A
36	125.5	3.5	785	12	US-10-369-493-5133	Sequence 5133, App
37	124.5	3.5	859	12	US-10-369-493-25	Sequence 25, App
38	124.5	3.5	859	12	US-10-369-493-251	Sequence 251, App
39	124.5	3.5	868	12	US-10-291-172-251	Sequence 106, App
40	123.5	3.5	868	9	US-09-800-729-106	Sequence 1635, A
41	122	3.4	710	12	US-10-369-493-1665	Sequence 1230, App
42	121	3.4	550	12	US-10-369-493-1230	Sequence 1168, App
43	119.5	3.4	538	12	US-10-369-493-1168	Sequence 1393, A
44	118.5	3.3	780	12	US-10-369-493-13393	Sequence 3, App
45	118.5	3.3	1781	9	US-09-738-877-3	

ALIGNMENTS

RESULT 1
US-09-907-907A-42
; Sequence 42, Application US/09907907A
; Publication No. US20030099660A1
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: GENES DISPLAYING ENHANCED EXPRESSION DURING CELLULAR SENESENCE A
; TITLE OF INVENTION: TERMINAL CELL DIFFERENTIATION AND USES THEREOF
; FILE REFERENCE: A34584-A-PCT-USA (070050.1664)
; CURRENT APPLICATION NUMBER: US/09/907, 907A
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: US 09/243,277
; PRIOR FILING DATE: 1999-02-02
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-907A-42

Query Match 100.0%; Score 3557; DB 11; Length 705;
Best Local Similarity 100.0%; Pred. No. 3.4e-302;
Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	DGFLLPRDRALTOLOVRA	MSSAGSAVA	VVDVDR	LTETSSG	KLAFADG	SNVVGSD	60
DB	1	DGFLLPRDRALTOLOVRA	MSSAGSAVA	VVDVDR	LTETSSG	KLAFADG	SNVVGSD	60
QY	61	TAVMTAASKRPSQ	EMPLVVDYR	OKAAAGRI	PTNYLR	REVGTSD	KEILTSRI	120
DB	61	TAVMTAASKRPSQ	EMPLVVDYR	OKAAAGRI	PTNYLR	REVGTSD	KEILTSRI	120
QY	121	IRLPFAGYFDY	TVLCNLA	AVDGN	EPDVA	INGASVAL	SLSDIP	180
DB	121	IRLPFAGYFDY	TVLCNLA	AVDGN	EPDVA	INGASVAL	SLSDIP	180

Qy	181	GEYVNPFRKMSSTTLVVAQAPKSGIWMLESAENIILQODFCHAKVGVKTTQOI	240
Db	181	GEYVNPFRKMSSTTLVVAQAPKSGIWMLESAENIILQODFCHAKVGVKTTQOI	240
Qy	241	GIQOLVKETGVTKRTPOKLFTPSPEIIVKTYHKLAMERLYAVFTDYEHDKVSDEAVNKIR	300
Db	241	GIQOLVKETGVTKRTPOKLFTPSPEIIVKTYHKLAMERLYAVFTDYEHDKVSDEAVNKIR	300
Qy	301	LDTEBOLKEKTPPEADPYEIIISFNVAKEVRSIYLVNKKRCDGRDLTSLNVSCEVUMF	360
Db	301	LDTEBOLKEKTPPEADPYEIIISFNVAKEVRSIYLVNKKRCDGRDLTSLNVSCEVUMF	360
Qy	361	KLHGSALEFORGQTOVLCTVTDFDSLSEGIKSDOVITAINGIKDXNFMLHYEPFPAATNEI	420
Db	361	KLHGSALEFORGQTOVLCTVTDFDSLSEGIKSDOVITAINGIKDXNFMLHYEPFPAATNEI	420
Qy	421	GKVTGLNRRLEIGHGALAKALYPIVPRDPPTIRTSSEVSNSSSSMAKSCGSLAMD	480
Db	421	GKVTGLNRRLEIGHGALAKALYPIVPRDPPTIRTSSEVSNSSSSMAKSCGSLAMD	480
Qy	481	SGVPISSAVAGVAILGVTKTDPBKGIEDYRLTDLIGIBDYNQDMPFKIAGTKNGITATL	540
Db	481	SGVPISSAVAGVAILGVTKTDPBKGIEDYRLTDLIGIBDYNQDMPFKIAGTKNGITATL	540
Qy	541	QADILKPEIPIKIWEALIQOASVAKKEILOIMNTKISKPRAKRENGGVVETTVQVPLSKR	600
Db	541	QADILKPEIPIKIWEALIQOASVAKKEILOIMNTKISKPRAKRENGGVVETTVQVPLSKR	600
Qy	601	AKFVPGGYNLKKLOAETGVTISQVDEETFSVFAPTPSVMEADRFIETICKDQEOOLE	660
Db	601	AKFVPGGYNLKKLOAETGVTISQVDEETFSVFAPTPSVMEADRFIETICKDQEOOLE	660
Qy	661	FGAVYTAITETIRDTGVNVKLYPNMTAVLLNHTQIDNERLNIILP	705
Db	661	FGAVYTAITETIRDTGVNVKLYPNMTAVLLNHTQIDNERLNIILP	705

RESULT 2
 US-09-907-907A-44
 Sequence 44, Application US/09907907A
 Publication No. US20030099660A1
 GENERAL INFORMATION:
 APPLICANT: Fisher, Paul B.
 APPLICANT: Leszczyńska, Magdalena
 TITLE OF INVENTION: GENES DISPLAYING ENHANCED EXPRESSION DURING CELLULAR SENESENCE
 TITLE OR INVENTION: TERMINAL CELL DIFFERENTIATION AND USES THEREOF
 FILE REFERENCE: A34584-A-PC1-USA (070050.1664)
 CURRENT APPLICATION NUMBER: US/09/907,907A
 CURRENT FILING DATE: 2001-07-16
 PRIOR APPLICATION NUMBER: US 09/243,277
 PRIOR FILING DATE: 1999-02-02
 NUMBER OF SEQ ID NOS: 51
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 44
 LENGTH: 705
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-907-907A-44

Query Match 100.0%; Score 3557; DB 11; Length 705;
 Best Local Similarity 100.0%; Pred. No. 3,4e-302;
 Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGFLLPDRDALQLOVRALMSSSGRAVVDVGNRLLETSSKILRAFAGSAVVOGSD 60
 |||
 DB 1 DGFLLPDRDALQLOVRALMSSSGRAVVDVGNRLLETSSKILRAFAGSAVVOGSD 60
 |||
 QY 61 TAVVAVTASKTSPSPSQFMPLVVDYRQKAAAGRIPTNYLRREVTSDEKITLSRIIDRS 120
 |||
 DB 61 TAVVAVTASKTSPSPSQFMPLVVDYRQKAAAGRIPTNYLRREVTSDEKITLSRIIDRS 120
 |||
 QY 121 IRPLPAGYFDTQYLCMLAVDGVNEPDTAINGASVALSLSDIPMNGPVGAVRIGIID 180
 |||

Dh	121	IRFLPAGFYDTQYLCNLMADVGNBPVLAINGASVALSLSDIPMNBPVGAVRGIIID	180
Qy	181	GBTVNPFTRKEMSSSTLNLVAGAPKSOJVMLEASANILQODFCHAIKVGKTYTOIIQ	240
Dh	181	GBYVNPTRKEMSSSTLNLVAGAPKSOJVMLEASANILQODFCHAIKVGKTYTOIIQ	240
Qy	241	GIQOLVKETGVYKRRPOKLFPSBPIVKYTHLAMERLYAVFTDYEHDKVSDAEANKIR	300
Dh	241	GIQOLVKETGVYKRRPOKLFPSBPIVKYTHLAMERLYAVFTDYEHDKVSDAEANKIR	300
Qy	301	LDTEBOLKBPPEADPYEIIIESFNVAKEVPSIYLANEYKRCQGRDJTSLRNVSCVDNF	360
Dh	301	LDTEBOLKBPPEADPYEIIIESFNVAKEVPSIYLANEYKRCQGRDJTSLRNVSCVDNF	360
Qy	361	KTLLGSAALFORQOTQVLCVTJPEDSLESGISKDQVITAINGIKDKPMFLHYBPYPVATNEI	420
Dh	361	KTLLGSAALFORQOTQVLCVTJPEDSLESGISKDQVITAINGIKDKPMFLHYBPYPVATNEI	420
Qy	421	GKVTGLNRELGHKMLAEKALYPVLPDPFPITIRYTSBYLESNGSSSMASACGSLALMD	480
Dh	421	GKVTGLNRELGHKMLAEKALYPVLPDPFPITIRYTSBYLESNGSSSMASACGSLALMD	480
Qy	481	SGVPISSSAVAGAIQLVMTKDEPKBIEBYRLTJLIIGIEDYNGMDPFIAGNKGITLAL	540
Dh	481	SGVPISSSAVAGAIQLVMTKDEPKBIEBYRLTJLIIGIEDYNGMDPFIAGNKGITLAL	540
Qy	541	QADITLPGIPIKIWEBAIQOASVAKKEILQINNKTIISKPRASKENGPAVETVOYPLSKR	600
Dh	541	QADITLPGIPIKIWEBAIQOASVAKKEILQINNKTIISKPRASKENGPAVETVOYPLSKR	600
Qy	601	AKFVPGGYNLKKLQALFTGVTTISQVDEETFSVPALTPSVYMEHARDPITIEICKDOEQOLE	660
Dh	601	AKFVPGGYNLKKLQALFTGVTTISQVDEETFSVPALTPSVYMEHARDPITIEICKDOEQOLE	660
Qy	661	FGAVVTATITTEIRDTGNVVKLYPNNMTAVULNHTQUDNERNLITLLP	705
Dh	661	FGAVVTATITTEIRDTGNVVKLYPNNMTAVULNHTQUDNERNLITLLP	705

```

/ RESULT 3
/ US-09-815-242-5443
/ Sequence 5443, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ TITLE OF INVENTION: Prokaryotes
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0

```

SEQ ID NO 5443
LENGTH: 696
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5443

Query Match 33.5%; Score 1191.5; DB 9; Length 696;
Best Local Similarity 39.2%; Pred. No. 4.1e-95;
Matches 274; Conservative 117; Mismatches 249; Indels 59; Gaps 12;

24 SAGRAVAVNDLGNRLKLEISSGKLARFADGSAVQSGDPAVMTAVSKTPSPQMPPLV 83
2 SEKKVFTKEMAGRSLLTETQLAKQANGAVLVYRGDVLVLTATASKEPRGDFPLTV 61
84 DYKQAAAGRIPTNYLREVEGSDKEILTSRIIDRSIRPLPAGFYDTQVLCNLAVD 143
62 NYEEKVYAAGKIPGFKRREGRGDATTALRIDRPIPLPKYKHDVQIMNVLSAD 121
144 GNEBDVLAINGASVALSLSDIPMNGPVGAVRIGIIDGVYVNPTRKEMSSSTLNLVAG 203
122 PDCSPQMAAMIGSSMALSVSDIPFGPIAGVNVGIDKYLINPVEKEVSRDLLEVAG 181
204 AKSQIWLLEASAEINLODFCHAIKVG-----VKYTOOIIQGIQOLVKEGVTKRP 256
182 HKDAVNVVEAGASITTEQEMLEAIFFGHEEIORLVDFQOQIVDHIOVKQE----- 232
257 OKLFTPS-----PEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEQKE 309
233 ---FPAERDEALVERIKSLTEBKGLKETVLF-----DKOORDENLNLK---BEIVNE 281
310 KPEENDP-----YEIISFNVAKEVRSIVLNEYKCDGRDLTSLANVSCVDMPKTL 363
282 FIDEBDPENELLKEVYALINELVKEVRRLLADEKIRPDGRKPDRIPLDSEVGLPRT 341
364 HGSALFORQOTQVLCVTPDSLSGKSDQVYTAINGIKDKNFMLYEFPYATNIEGKV 423
342 HSGGLFTRGQYALSVLTIGAL-----GDYQOLIDGPEBEKRFMHYVNPFSVGETGPV 397
424 TGLNRELGALAEKALVPIV--RDPFTIRVTSVLESNGSSSMASACGSLALMDS 481
398 RAPGRREIGHALGERALKYIIPDADPFYTRIVSEVLESNGSSQASICSTLALMDA 457
482 GVPISAVAGVALGLVTKTDPKGEIEYRLTLIDIGEDYNGDMDFKAGTNGKITLQ 541
458 GVPKAPVAGIAGLVTRD-----SYTLITDIQMEDALGDMDFKAGTNGKITLQ 510
542 ADIKLPGIPKIVMEAIQOASVAKKEILOINMKTISKPRARKENGVPVETVQVLSKRA 601
511 MDIKIDGLTRRIIEBALQARRGRLEIMNMHLOTIDQRTLSAVAPKVVMTIKPKDIR 570
602 KRVGSGYNLKKLOAETGVTTISQVDEETSVFAPTPSVHMEARDPTEICKDDQEQLEF 661
571 DVIGGKGINIEIDETGVKLDIEODGTIFGAVDQAMINRAREIIEITR-----EAEV 625
662 GAVYATITIRDTGVMMVLYPNMTAVLHNTQLDNERL 700
626 GQTYQATVRIEKYGAFLVPGKOA-LAHISQISKNRI 663

RESULT 4

US-09-815-242-12325
Sequence 12325, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Hasebeck, Robert
APPLICANT: Ohlsen, Karl U.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12325
LENGTH: 698
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12325

Query Match 33.5%; Score 1191.5; DB 9; Length 698;
Best Local Similarity 39.2%; Pred. No. 4.1e-95;
Matches 274; Conservative 117; Mismatches 249; Indels 59; Gaps 12;

24 SAGRAVAVNDLGNRLKLEISSGKLARFADGSAVQSGDPAVMTAVSKTPSPQMPPLV 83
2 SEKKVFTKEMAGRSLLTETQLAKQANGAVLVYRGDVLVLTATASKEPRGDFPLTV 61
84 DYKQAAAGRIPTNYLREVEGSDKEILTSRIIDRSIRPLPAGFYDTQVLCNLAVD 143
62 NYEEKVYAAGKIPGFKRREGRGDATTALRIDRPIPLPKYKHDVQIMNVLSAD 121
144 GNEBDVLAINGASVALSLSDIPMNGPVGAVRIGIIDGVYVNPTRKEMSSSTLNLVAG 203
122 PDCSPQMAAMIGSSMALSVSDIPFGPIAGVNVGIDKYLINPVEKEVSRDLLEVAG 181
204 AKSQIWLLEASAEINLODFCHAIKVG-----VKYTOOIIQGIQOLVKEGVTKRP 256
182 HKDAVNVVEAGASITTEQEMLEAIFFGHEEIORLVDFQOQIVDHIOVKQE----- 232
257 OKLFTPS-----PEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEQKE 309
233 ---FPAERDEALVERIKSLTEBKGLKETVLF-----DKOORDENLNLK---BEIVNE 281
310 KPEENDP-----YEIISFNVAKEVRSIVLNEYKCDGRDLTSLANVSCVDMPKTL 363
282 FIDEBDPENELLKEVYALINELVKEVRRLLADEKIRPDGRKPDRIPLDSEVGLPRT 341
364 HGSALFORQOTQVLCVTPDSLSGKSDQVYTAINGIKDKNFMLYEFPYATNIEGKV 423
342 HSGGLFTRGQYALSVLTIGAL-----GDYQOLIDGPEBEKRFMHYVNPFSVGETGPV 397
424 TGLNRELGALAEKALVPIV--RDPFTIRVTSVLESNGSSSMASACGSLALMDS 481
398 RAPGRREIGHALGERALKYIIPDADPFYTRIVSEVLESNGSSQASICSTLALMDA 457
482 GVPISAVAGVALGLVTKTDPKGEIEYRLTLIDIGEDYNGDMDFKAGTNGKITLQ 541
458 GVPKAPVAGIAGLVTRD-----SYTLITDIQMEDALGDMDFKAGTNGKITLQ 510
542 ADIKLPGIPKIVMEAIQOASVAKKEILOINMKTISKPRARKENGVPVETVQVLSKRA 601
511 MDIKIDGLTRRIIEBALQARRGRLEIMNMHLOTIDQRTLSAVAPKVVMTIKPKDIR 570
602 KRVGSGYNLKKLOAETGVTTISQVDEETSVFAPTPSVHMEARDPTEICKDDQEQLEF 661
571 DVIGGKGINIEIDETGVKLDIEODGTIFGAVDQAMINRAREIIEITR-----EAEV 625

Qy 662 GAVTATITIRDTGVWVKLYPNMTAVLLHNTOLDNERL 700
Db 626 GQYQATVKRIEKYGFVGLFGKDA-LHHISQISKRI 663

RESULT 5

US-09-815-242-13768
Sequence 13768, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: EITRA.011a
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23,207,727
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27,931
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: PASTESEQ for Windows Version 4.0
SEQ ID NO 13768
LENGTH: 721
TYPE: PRT
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(721)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13768

Query Match 32.6%; Score 1158; DB 9; Length 721;
Best Local Similarity 39.9%; Pred. No. 3.7e-92;
Matches 270; Conservative 117; Mismatches 250; Indels 40; Gaps 11;

Qy 35 GNRKLEISSGKLAFADGSAVVGSDTAVMTAVSKTPSPSQ-FMPLVVDYRQKAAAG 93
Db 22 GQHXTLETGMARQATAAVWVSMDDTAVFTVVGQKAKGQDFPFLTVVYQERTVAG 81
Qy 94 RIFPNLYRREVGTSDKEILTSRIIDRSIRPLFPAGFYDTQVLCNLAVDGVNEPDLAI 153
Db 82 RIPSFXRREGRSPSGETLIARLLIDRPVRPLFPEGFVNEVVIATVVSQVNPVDIYAM 141
Qy 154 NGASVALSLSDIPNKGPGAVRIGIIDEVYVNPVRKMSSTINLVVAGAPKQIWL 213
Db 142 IGASAAALSLSGIPENGPAGARVGITNDQVYVNPQDELKSKLDLVVAGT-EAAVLNVE 200
Qy 214 ASAENILQODFCCHAIKVGKTYQOIIQGIQOLVKEVTKRTPQKLTFTSPSEIYK----- 268
Db 201 SEALLSDPTMLGAVFHEGQGVVIAINDLVKAGSPRMDQ-----PEAVNDALNA 254
Qy 269 YTHKLAMERL--YAVFTDYHDKVSDEAVNKIRLDTEDQKKEFPADPEYIEISFNV 325
Db 255 RVALLAESRLSDAVRI-----TDKQERYAVQVDVIKSEITIEQLADEDTLDANEIGETIHA 309

Qy 326 VAKVFSIVLNEKXRCDCGRLTSLRVNVSCEVDMFKTLHGSALFORQOTVLCVTFDSL 385
Db 310 IERNVRSRLVLAGEPRIIDGHEKMIKRLDVRTGVLPPTHSSALFTRGETQALVTATIGT- 368
Qy 386 ESGIKSDQVITAINGIKDKNFMHYEPFPYATNEIGKVTGLANRELGHALAEKALYPIV 445
Db 369 ---ARDAQVLDLMGERTDSFLPFYNNPPYSVGETGVGSPKREIGHGLAKRGVLAVM 425
Qy 446 P--RDPEPTIRYVSEVLESNGSSSMASACGSLALMSGVPISSAVGVAILGYTKDPE 503
Db 426 PDMKFPYTVRVAVEITSENGSSSMASVCSASIALMDAGVPIKAAVAGIAMGLVKEED-- 483
Qy 504 KGEIEDRYLTDIIGIDYNGDMDFKIAGNNKGTALQADIKLPGIPIKIWMENIQOASV 563
Db 484 ----NYVLSDILGDDHDMDFKAGSRDGSALQNMKIKETIKELINQVALNDAKG 538
Qy 564 AKKEILQIMNKTISKPRASRKNQPVVETVQVPLSKRAKVGPGVNLKRLQAEVGTIS 623
Db 539 ARLLHIGVMOAIVAPRGDISFPAPRIHTIKISTDKIKDVIKGGGSVIRALTEETGTIE 598
Qy 624 QVDEETSVFAPRPSVNHARDFTIEICKDQEQLEFGAVYATATIEIDTGMVLYP 683
Db 599 IEDDGYKIAATDGEKAKVARIIEI-----TAEIEVGRITVNSKVTRIVDFGAFVAGG 653
Qy 684 NMTAVLLHNTOLDNERL 700
Db 654 GKSG-LVHISQIADKRV 669

RESULT 6

US-09-815-242-10312
Sequence 10312, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: EITRA.011a
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23,207,727
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27,931
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: PASTESEQ for Windows Version 4.0
SEQ ID NO 10312
LENGTH: 734
TYPE: PRT
ORGANISM: Escherichia coli
US-09-815-242-10312

Query Match 32.5%; Score 1155; DB 9; Length 734;
Best Local Similarity 39.9%; Pred. No. 6.3e-92;
Matches 269; Conservative 116; Mismatches 252; Indels 36; Gaps 11;

QY 35 GNRKLEISSGKLARFADGSAVVGSDTAVMTAVSKTKSPSQ-FMPLVVDYROKAAAG 93
DB 35 GQHTVLETGMAARQATAVAVMSMDTAVFVTVGQKAKPGDFFPLTVNQGERTYANG 94
QY 94 RAPTNTLRREVTSDEKILTSRIIDSIRPLFPAGFYDTQVLCNLAAVDGNEBVLAI 153
DB 95 RLPGSFRRRGRSEBETLIRLIDRIRPLFPFGGVNEQVATVAVNPOVNPDIIVAM 154
QY 154 NGASVALSLSDIPMNGPVGAVRIGIIDGEYVNPTRKEMSSSTLNVAGAPSOIWMLE 213
DB 155 IGASVALSLSDIPMNGPVGAVRIGIIDGEYVNPTRKEMSSSTLNVAGAPSOIWMLE 213
QY 214 ASANILLODPCHAIKIVGVYTOOIIQIIOOLVETGVTKEPFLPQKLFPSPEIYVYIP- 270
DB 214 SEAGLLESDQMLGAVVFGHQOQVAVIQLNINELVKEG-----KPRMWOGEPPVNEALNAV 269
QY 271 HKLAMERL---YAVFPDYEHDKVSRDBAVNKRRLDTEBOLKEKPEADPYEIESFNVA 327
DB 270 AALAEARLSDAVRI-----TDKQERYAOVDVISEFTIATLLADEFTLDENELGEILHAIE 324
QY 328 KEVFRSIVLNEYKRCGRDLTSLRNVSCEVDMKTLHGSALPQRCOTQVLCVTPDSLES 387
DB 325 KAVVMSRVLAGERIDGRKMDIKGLDVRTGVLPRTHGSALFTRGTOALVTAITGT--- 381
QY 388 GIKSDOVITAINIGIKDKNFMHYEPFPYATNEIGKVTGLNRRELGHGALAERALYVIP- 446
DB 382 -ARDQVLDLGMERIDTFLFHNFPYSVGETGVNGVSPKREIGHGRALAKXGLAVAMD 440
QY 447 -RDPFTTIVTSVLESNGSSSMASACGSLALMDSGVPISSAVAGVAILGVTKDPEKG 505
DB 441 MDKFPYVAVSVSEITSSNGSSSMASACGSLALMDSGVPISSAVAGVAILGVTKDPEKG 496
QY 506 BIEDRYLTDIIGIEBYNDGMDPKIAGTNGKGTALQADIKLPGIPIKIYMEALQOASVAK 565
DB 497 ---NYVLSDBILDEDEHLDGMDPKVAGSDGSLALMDIKIGITKEIMQVLANQKGR 553
QY 566 KEILOIMNKTISPRASRKNQGVVETVQVPLSKRAKFPVGGVYNKKLOAETGVYISQV 625
DB 554 LHLIGMEQAINAPRODISEFAPRITIKINPKIKDVIGKGSVIRALTEETGTTIEE 613
QY 626 DEETSVPAPTPSVMEADFTIEICKDOEQLEFAGVYATITIEIRDTGVVVKLYPNM 685
DB 614 DQGVIAATIDDEKAKHARIRIEI-----TAIEIGVYTVGTVRIVDFGAFAVIGGK 668
QY 686 TAVLHNTOLDNERL 700
DB 669 EG-LVHISQIADKRV 682

PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10989
LENGTH: 709
TYPE: PRY
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(709)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-10989

Query Match 31.9%; Score 1134; DB 9; Length 709;
Best Local Similarity 36.8%; Freq. No. 4.5e-90;
Matches 249; Conservative 137; Mismatches 251; Indels 40; Gaps 12;

QY 35 GNRKLEISSGKLARFADGSAVVGSDTAVMTAVSKTKSPSQ-FMPLVVDYROKAAAG 93
DB 11 GQHTVLETGMAARQATAVAVMSMDTAVFVTVGQKAKPGDFFPLTVNQGERTYANG 70
QY 94 RAPTNTLRREVTSDEKILTSRIIDSIRPLFPAGFYDTQVLCNLAAVDGNEBVLAI 153
DB 71 KIPGGSFRRRGRSEBETLIRLIDRIRPLFPFGGVNEQVATVAVNPOVNPDIIVAM 130
QY 154 NGASVALSLSDIPMNGPVGAVRIGIIDGEYVNPTRKEMSSSTLNVAGAPSOIWMLE 213
DB 131 IGASVALSLSDIPMNGPVGAVRIGIIDGEYVNPTRKEMSSSTLNVAGAPSOIWMLE 189
QY 214 ASANILLODPCHAIKIVGVYTOOIIQIIOOLVETGVTKEPFLPQKLFPSPEIYVYIP- 270
DB 190 SEADLITBEOQLAAVVFQHQOQVAVIQLNINELVKEG-----KPRMWOGEPPVNEALNAV 245
QY 271 HKLAMERL---YAVFPDYEHDKVSRDBAVNKRRLDTEBOLKEKPEADPYEIESFNVAKEV 330
DB 246 KALAEARLSDAVRI-----TDKQERYAOVDVISEFTIATLLADEFTLDENELGEILHAIE 303
QY 331 FRSIVLNEYKRCGRDLTSLRNVSCEVDMKTLHGSALPQRCOTQVLCVTPDSLES 390
DB 304 VRGRIIAGEPRIDRVTQVTRALDICTGVLPRTGSAIFTRGTOALVTAITGT--- 358
QY 391 SD-QVITAINIGIKDKNFMHYEPFPYATNEIGKVTGLNRRELGHGALAERALYVIP--R 447
DB 359 RDAQIIDELETERDHFLEHNFPYSVGETGVNGVSPKREIGHGRALAKXGLAVAMD 418
QY 448 DEFTTIVTSVLESNGSSSMASACGSLALMDSGVPISSAVAGVAILGVTKDPEKEI 507
DB 419 EFPYVAVSVSEITSSNGSSSMASACGSLALMDSGVPISSAVAGVAILGVTKDPEKEI 471
QY 508 EDYRLTDIIGIEBYNDGMDPKIAGTNGKGTALQADIKLPGIPIKIYMEALQOASVAKKE 567
DB 472 EKFWLSDILDEDEHLDGMDPKVAGSDGSLALMDIKIGITKEIMQVLANQKGR 531
QY 568 ILOIMNKTISPRASRKNQGVVETVQVPLSKRAKFPVGGVYNKKLOAETGVYISQVDE 627
DB 532 ILGMEQAINAPRODISEFAPRITIKINPKIKDVIGKGSVIRALTEETGTTIEE 591
QY 628 ETPSVFA---PTPSVMEADFTIEICKDOEQLEFAGVYATITIEIRDTGVVVKLYP 683
DB 592 GTVIAAVDSVAAGVAGRIEIVAEV-----EAGVYIKKQVTRLDVFGAFAVIG 642
QY 684 NMTAVLHNTOLDNERL 700
DB 643 NKEG-LVHISQIADKRV 658

RESULT 8
US-09-907A-43
Sequence 43, Application US/09907907A
Publication No. US20030099660A1
GENERAL INFORMATION:
APPLICANT: Fisher, Paul B.
APPLICANT: Leszcyniecka, Magdalena
TITLE OF INVENTION: GENES DISPLAYING ENHANCED EXPRESSION DURING CELLULAR SENESENCE A
TITLE OF INVENTION: TERMINAL CELL DIFFERENTIATION AND USES THEREOF
FILE REFERENCE: A34584-A-PCT-USA (070050.1664)
CURRENT APPLICATION NUMBER: US/09/907, 907A
PRIOR FILING DATE: 2001-07-16
PRIOR APPLICATION NUMBER: US 09/243,277
PRIOR FILING DATE: 1999-02-02
NUMBER OF SEQ ID NOS: 51
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 43
LENGTH: 705
TYPE: PRT
ORGANISM: Bacillus subtilis
US-09-907A-43

Query Match 31.8%; Score 1132; DB 11; Length 705;
Best Local Similarity 37.4%; Pred. No. 6,7e-90;
Matches 251; Conservative 144; Mismatches 253; Indels 24; Gaps 10;

32 VDLGRKLEISSGLARADGSAVVGSDTAVMYAVSKTPSPQFPLVVDYRQKAA 91
10 IDWAGRTLVETGQAKQANGAVMIRYGDVAVLSTATSKBPKPLDFEPLTVNEERLYA 69
92 AGRIPTNLRREVSTGDEKEILTSRIIDRSIRPLFPAGYFVYTVQVLCNLAAVGVNEPVL 151
70 VKDIGGKIKKRGRESEKAVLASRLIDRPIRLFDGGRNREVQVSIYMSVDQNSSEMA 129
152 AINGASVALSLSDIPWNGPVGAVRIGIIDGEVYVNPTRKSSSTLNLVVAAPRSQITW 211
130 AMFGSLALSVSDIPFEGPIAGVTGRIDDPGFIINPTVQLEKSDINLVVAGT-KDAIIM 188
212 LEASGENTLODFCHAIKVKYKTYQOIIQGIQOLVKEGVTKTKQKLFSPSEIVKTXH 271
189 VEGAGDEVEPEEIMLAIHFGHEIRRLIAFOEIVAAVG-KEKSKIKFEIDEELINEVK 247
272 KLAMRLVAVFTDYEDHKVSRDEAVNKIRLDTESQLEKEKPEADRY-ELISSFNVAKEV 330
248 ALAEEDLKAIV--QVHEKGAARDALINEKNAVVAPEDEBHDITIKYKQLSKLVNE 305
331 FRSIIVNEKRCGDRDLSLRNVSCEDVMFKTLHGSALEFQRGQVLCVTVPDSIESGIK 390
306 VRRLITEKVRPDGRGVQIRPLSEVGLPRTHGSGLFTRGQTOALSVCITGAL---G 361
331 SDQVTAINGIKDKKFMILHYEPFPAATNEIGKVTGLANRELGHGLAKALYVPIR-RD 448
362 DVQIIDGIGVESKKEFMHYNPPQSVGETGPMRGPGRREIGHGLGRALPEVPIPSKD 421
449 PFETIRATSEVLESNGSSMASAGSIALMDSGVPISSAVAGVAIGVTKDPEKGEIE 508
422 FPYTRIVSEVLESNGSSQASICASTLAMDAIPITAPVAGIMGLV-----KSG--E 474
509 DYRLITDILIGIEDYNGDMFKIATNGKITALQADIKLPGIPIKIWEALIQASVAKKEI 568
475 HYTVLTIDIGMEDALGDMDFKVAGTEKGVTAILOMIKIEGLSRILIEBALQOAKKGRMEI 534
569 LOIMNKITSKPAASKENGPVETQVPLSKRAKVVGGGVNKLQAEFTQTTISQVDEE 628
535 LNSMATLSESRKELSRVAPKILITMTINDKIRDIVISGQKINKIIBETGVKIDIEODG 594
629 TFSVAPFPPSVVHBAARDFTIEICXPDQEOLEFGAVVYATITIEIDTGVVNLVYPMNTAV 688
555 TIFISTBESGQAKKII-----EDLVAREVNGQULYKAVRIKFGAFVETIFSGKSG- 648
689 LIHNTQDNRL 700

Db 649 LVHISLALERV 660

RESULT 9
US-09-815-242-10882
Sequence 10882, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10882
LENGTH: 704
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-815-242-10882

Query Match 31.7%; Score 1128.5; DB 9; Length 704;
Best Local Similarity 36.7%; Pred. No. 1,4e-89;
Matches 258; Conservative 147; Mismatches 245; Indels 53; Gaps 15;

13 LTQQLV-PALMSAGSAVAVDLGNKLEISSGLARADGSAVVGSDTAVMYAVSKT 71
1 MTEKQVFKTW-----GGRPLEVEIGQAKQANGAVVRYGDTVVLSSAAVASK 48
72 KPSSQSGMPPLVVDYRQKAAAGRIPTNLRREVSTGDEKEILTSRIIDRSIRPLFPAGYF 131
49 EAKVDPEPLTVNEEKMYAVGKIPGFTIKRGPRSSRATLTLIRPIRPMSEGRN 108
132 DTQVLCNLAAVGVNEPDLVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEVYVNPTRKE 191
109 EVQITNIVMSGEQCTBMAAMFGSSIALAISDIPFGPIAGVDVGRINGEVVLPVTEQ 168
192 MSSSTLNLVVAAPRSQITWLEASAEENILODFCHAIKVKYKTYQOIIQGIQOLVKEGV 251
169 AEQDIDIELTVAGT-KEAINVWESGAKVSEBDMIGALLFGDAIKELVAFQEEIVAAVVG 227
252 TKRPQKLFPSPEPIVATYTKHLAERYL-----AVFTDYEDKVS RDEAVNKIRLDTBEQ 306
228 PKMDVLL-----QVDADLKEEIPDAYNTMTKTAVMTE--EKAREVEIDKVQDYKEV 279
307 LKEKPE-ADPEYELIESEFNVA-----KEVRSIVLNEKRCGDRDLSLRNVSCEDVMFK 361
280 YAEKFAHEBAQULKVKQAELEKQVARELITIKIRPDGKLBIEIRLSSEVGLP 339
362 TLHGSALEFQRGQVLCVTVPDSIESGIKSDQVTAINGIKD-KFMILHYEPFPAATNEI 420

Db 340 RHVSGSLFTRGOTQALSVCTL---APLGEHQIIDL-GVODSKRFIHHYFPQPSVSGT 394
 Qy 421 GYVTLNRELGHGALAEKALYVIP--RDPFTIRVTSEVLESNGSSSMASACGSLAL 478
 Db 395 GBAAGRGREIGHGALGELAQAIIIPSEDPFTIRLVAELESNGSSSQASICAGTLAL 454
 Qy 479 MDGVPISAVAGVALGVLTKTDPKGEIIEYRLTLTGIEDYNGDMFKIAGTNKGIT 538
 Db 455 MDAGVPIKAPVAGIANGLVS-----DGENYTLITLDIGLEDHIGDMDFKVAAGTMDGIT 507
 Qy 539 ALQADIKLPGIPIKIVMEALIQASVAKKILQIMNKTIKPPASRKNQPVETVQVPLS 598
 Db 508 ALQMDIKIQGITEQILTEALDQAKRMBELIETTLTIAAPREBSQVAPKITEMIQKPA 567
 Qy 599 KKAQFVPGGYNLKKLQAGETGVTISQVDETFSPVAPFSPVMEHARDPTEICXDQDEQ 658
 Db 568 KIKVDYIGKGGETINSIIDETGVKIDIDQDGNVSIASSDAEMKKAKIIEBLTK-----E 622
 Qy 659 LEFGAVYATITTEIRDTGVVVKLYPNMTAVLLHNTQDNERLN 701
 Db 623 VEVGVYLAQVARIIEKFGAFVNLKIKGKG-LIHISQLANERYN 664

RESULT 10

US-09-815-242-4975
 ; Sequence 4975, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO: 4975
 ; LENGTH: 702
 ; TYPE: PRT
 ; ORGANISM: Enterococcus faecalis
 US-09-815-242-4975

Query Match 31.7%; Score 1127; DB 9; Length 702;
 Best Local Similarity 37.4%; Pred. No. 1,8e-89;
 Matches 254; Conservative 144; Mismatches 242; Indels 40; Gaps 13;
 Qy 35 GNRKLEISSGLARFADGSAVVGDTAVMTAVSTKSPSQFMPVLVDYRQKAAAGR 94
 Db 10 GGRPLEVEIGQLAKQANGAVLVRYGDTVVLAAVASKAKVDVFPFLVYNEEKMTAVAK 69
 Qy 95 IPTNLRREVGTSDEKILTSRIIDSIRPLFPAGYFYDTQVLCNLAVDGVNEPVLAIN 154

Db 70 IPGFGIKRGRPSERATILTAALIDRPIRMPSBGRFNEVOITNTIVMSVEODCTPEMAAMF 129
 Qy 155 GASVALSLSDIPWNGPVGAVRIGIIDGEYVNPTRKEMSSSTLNLVAGAPSOIWMLEA 214
 Db 130 GSSLALASDIPFDDPIGAVGVRINGSEYVNLPEYQAEQOTIDELTAVGT-KEALNMVES 188
 Qy 215 SAENILQODPCHALKVGVYTOQIIQGIQOLVKETGVKRTPOKLFPTSPBIVKYTHKLA 274
 Db 189 GAKVSEBDMGALLFGFDALKELVAPQEBELVAANGKRMVDLL-----QVDADLKEEI 243
 Qy 275 MERLY-----AVTDYEDHKVSRDEAVNKIRLDEEQLKEKPE-ADPEYIEISFNVA- 327
 Db 244 FDAAVNTKTKVMTB---EKLAREVEIDKVDQVAVAEKFAHEBEBAQLKEVQOIAE 300
 Qy 328 ---KEVPSIYLANEYKRCDDGDLTSLRVSGEVMFKTLHGSALPQROTQVLCVTPDS 384
 Db 301 DLEKQVRELTLTIDIRPDGKRLDEIRLSEVSTLPPVHSGSLFTRGOTQALSVCTL-- 358
 Qy 385 LESGIKSDQVITAINGIRD-KNFMILHYEPFYATNIEIGVTGLNRELGHGALAEKALY 443
 Db 359 --APLGEHQIIDL-GVODSKRFIHHYFPQPSVSGTGRASPGREIGHGALGELAQA 415
 Qy 444 VIP--RDPFTIRVTSEVLESNGSSSMASACGSLALMDSGVPISSAVAGVALGVLTKTD 501
 Db 416 IIPSEDEPPTIRLVAEVLSENGSSSQASICAGTLALDADGVPIKAPVAGIANGLVS--- 472
 Qy 502 PEKGEIEYRLTLTGIEDYNGDMDFKIAGTNNGITLQMDIKLPGIPIKIVMEALIOA 561
 Db 473 ---DGENYTLITLDIGLEDHIGDMDFKVAAGTMDGITLQMDIKIQGITEQILTEALDQA 528
 Qy 562 SVAKKEILQIMNKTIKPPASRKNQPVETVQVPLSRAPFGGYNLKKLQAGETGVT 621
 Db 529 KKAEMELIETTLTIAAPREBSQVAPKITEMIQKPAKIKOVIGGSETINSIIDETGVK 588
 Qy 622 ISQVDETFSPVAPFSPVMEHARDPTEICXDQDEQLEPAVYATITTEIRDTGVVVKL 681
 Db 589 IDIDQDGNVSIASSDAEMKKAKIIEBLTK-----EVEGVYLAQVARIIEKFGAFVNL 643
 Qy 682 YPNMTAVLLHNTQDNERLN 701
 Db 644 IKGKG-LIHISQLANERYN 662

RESULT 11

US-09-815-242-11755
 ; Sequence 11755, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 11755
LENGTH: 613
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-815-242-11755

Query Match 30.8%; Score 1097; DB 9; Length 613;
Best Local Similarity 40.4%; Pred. No. 6,2e-87;
Matches 248; Conservative 110; Mismatches 222; Indels 34; Gaps 9;

QY 39 LEISGKLARFADSGAVVQSGDTAVMTAVATKSPSPQ-FMPLVVDYRQKAAARPT 97
DB 1 MLETKMARQATAVVMSMODTAVFVVGQKAKKQDPFLVAVQETVYAGKIPG 60
QY 98 NYLRREVQTSDEKILTSRIIDRSIRPLFPAGYFYDTQVLCNLAVDGVNEPDLAINAS 157
DB 61 GFFRREGSRSGETLIARLIDRPRPLFPBGFAVEQVIAIVSVNPNVDIYAMIGAS 120
QY 158 VALSLDIPMNGPVGAVRIGIIDGYVNPFRKENSSTLNLVAGAPKSOIVMLASAE 217
DB 121 AALSLGIPFNGPICAARVGYINDQVYVNPQEBLKSRLVAGT-EAALVAVESAE 179
QY 218 NILQDPFHAKVGYKTYQOIIQGIQOLVKETGYTKRPQKLPFSPPEIVK-----YTHK 272
DB 180 LLSQKMGAVVFGHQOQIVIQINDLVKAGKPRMDQ-----PAAVDEALNARVAA 233
QY 273 LAMERL---YAVFTDYEDHKVSRDEAVNKIRLDEEQLEKEFPPEADPEIIESFVNAKE 329
DB 234 LAERLSDAVRI-----TDQERYAQVUVISETIATVADEDTLDANLEGEIILHAIEKN 288
QY 330 VFRSIVLEVYRCQGRDLTSLRNVSCVDMFKTLHGSALFORQOTVLCVTFPSLESIGI 389
DB 289 VFRSIVLEVYRCQGRDLTSLRNVSCVDMFKTLHGSALFORQOTVLCVTFPSLESIGI 389
QY 390 KSDQVITANGIKDKNFMHYEPFPYATNEIGKVTGLNRRELGHGALAEKALYPIR-R 447
DB 349 NIDELM---GERDTSFLPHNFPQYISVGEQMGSPRRREIGHRLAKRGLVAMPTE 404
QY 448 DPEFTIRVTSVLESNGSSSMASACGSLALMDSGVPISSAVAGVAILGVTKEPKEGI 507
DB 405 EFTPTVRVVSITSSNGSSSMASVCGSLALMDSGVPISSAVAGVAILGVTKEPKEGI 507
QY 508 EDYRLLDLIGIEDYNGMDPKIAGTNGKIGTALQADIDLPGLPIKIVMEAIQOASVAKKE 567
DB 459 -NFWVLSLIDGDEHLDGMDPKVAGSRDQISALQMDIKIEGITYKEIMQVAINQAKGAILH 517
QY 568 ILQIMNKTIISKPRARKENGVPVETVOVPLSKRAKFGVPGVNLKKQLAETGVITISQVDE 627
DB 518 ILGVNEQAINAPRGDISFAPRIHTIKINPKIKQVIGKGSVIRALTEETGTTIEEDD 577
QY 628 ETPSVFAPTPSVMH 641
DB 578 GTVAKIATDGDKAH 591

RESULT 12
US-09-815-242-12050
Sequence 12050, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Tremick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21, 078
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 12050
LENGTH: 701
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12050

Query Match 30.8%; Score 1097; DB 9; Length 701;
Best Local Similarity 36.5%; Pred. No. 7,7e-87;
Matches 248; Conservative 141; Mismatches 250; Indels 40; Gaps 14;

QY 35 GNRLEISGKLARFADSGAVVQSGDTAVMTAVATKSPSPQ-FMPLVVDYRQKAAAG 93
DB 11 GQSTVLTETRIARQAGAVLVMTMDVSLVTVVGAASPAEGBDFPLSHYQKTVAAAG 70
QY 94 RIPFNVYLRREVQTSDEKILTSRIIDRSIRPLFPAGYFYDTQVLCNLAVDGVNEPDLAI 153
DB 71 RIPGFFRREGSRSGETLIARLIDRPRPLFPBGFAVEQVIAIVSVNPNVDIYAMIGAS 130
QY 154 NGASVALSLDIPMNGPVGAVRIGI-IDGEVNPFRKENSSTLNLVAGAPKSOIVML 212
DB 131 IGTSALAIISGIPFAGTGAARVGFPEIGIILNPTEQOASSLDDVAVAGT-EDAVLVAV 189
QY 213 EASAKENIQDPFHAKVGYKTYQOIIQGIQOLVKETGYTKRPQKLPFSPPEIVKY 269
DB 190 ESEADELTDQMGAVLFAHDEPQAVIRAVKELAAEKGKAMPDKAPAE-NTVLVNAIKA 248
QY 270 THKLAMERLYAVFTDYEDHKVSRDEAVNKIRLDEEQLEKEFPPEADPEIIESFVNAKE 324
DB 249 ELGSAISQAVYTI-TIKQDYNNRIGELRDOAVLFAEEB---GKFPAS---EVKQVFG 299
QY 325 VVAKEVRSIVLEVYRCQGRDLTSLRNVSCVDMFKTLHGSALFORQOTVLCVTFPSLESIGI 384
DB 300 LLERTYRENTVNGKPIIDGDRTRVAPLRIEYGVLTGKHSALFTGSETQALVAVATLGT 359
QY 385 LESGKSDQVITANGIKDKNFMHYEPFPYATNEIGKVTGLNRRELGHGALAEKALYPIV 444
DB 360 ---ARDAQLDLTLEGRKDAFPMHYNFPFVSGECGRMSPGRRREIGHRLARAGVAA 415
QY 445 IPR--DPEFTIRVTSVLESNGSSSMASACGSLALMDSGVPISSAVAGVAILGVTKEPKEGI 502
DB 416 LPTQDEPPTIRVVSITSSNGSSSMASVCGASLALMDSGVPISSAVAGVAILGVTKEPKEGI 502
QY 503 EKGIEDYRLTDLIGIEDYNGMDPKIAGTNGKIGTALQADIDLPGLPIKIVMEAIQOASVAKKE 562
DB 472 ---EGKFAVLTDLIGDEHLDGMDPKVAGSRDQISALQMDIKIEGITYKEIMQVAINQAKGAILH 517
QY 563 VAKKEILQIMNKTIISKPRARKENGVPVETVOVPLSKRAKFGVPGVNLKKQLAETGVITISQVDE 622
DB 529 EARLNLIGQMNQVIAKRAELSENAFPMLOMKIDSDKIRIVIGKGAITINGICETASII 588
QY 623 SQVDEFTSVFAPTPSVMEARDFTEICKDQEQLEFQAVYATITTEIRDTGVMKLY 682
DB 589 DIEDDGSVKIYGTETKAAEAKLRLVLAI-----TAAEIKIYGVKVERIVDFGAFVNL 643

Qy 683 PMTAVLHNTOLDNERLN 701
Db 644 PKDGG-LVHISQISDKRID 661

RESULT 13

US-09-815-242-13699
Sequence 13699, Application US/09815242
Patent No. US020061569A1
GENERAL INFORMATION:
APPLICANT: Haseibeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13699
LENGTH: 737
TYPE: PRF
ORGANISM: Streptococcus pneumoniae
US-09-815-242-13699

Query Match 29.6%; Score 1054; DB 9; Length 737;
Best Local Similarity 34.6%; Pred. No. 4,9e-83;
Matches 234; Conservative 148; Mismatches 257; Indels 38; Gaps 9;
Qy 37 RLLEISGGLARFADGSAVVGSDTAVMTAVSKTKPSQMPPLVVDYRQRAAAGRIP 96
Db 13 RELIVETGVAAQANGSVVVRGSESTVLAAMSKKMAAGDFFPLQVNTBEKRYAAGKFP 72
Qy 97 TNYLREVSTSKELIIRISIRPLFPAGFYDVOYLCMLAVDGVNEPDVLAINGA 156
Db 73 GGFMRGEGSPDATTALIDRIPRMAEGFRNEVOYINVLSTIDENASAPMAAFMS 132
Qy 157 SVALSLSIDIPMNGPVAGVARIIGIIGERYVNPTRKENSSTLNLVAGAPKSOIWMLEASA 216
Db 133 SIALSISIDIPFDGPIAGVGVYDQGIINPSQEQALQELLETVAAGT-KHAINNVESSGA 191
Qy 217 ENILLODFGHAIVKQVKTQOIIIGIQQLVYKGTGVRPQQLFTPSPPIVYKTHLAME 276
Db 192 KLSLSIMLEALKGAEVAKELIAFQSEKIVAAVNGEK-----AEVELHVVDELQAE 243
Qy 277 RLVAFTVDE-----HDKVSDEAVANKRIIDTEBQLEKFPADPF-----EIISEFNV 326
Db 244 IIAAVNSDIQXAVOVEKELAREATQAVDQYTAVEEKKANHEEDRIMRDVAELIEM 303
Qy 327 AAEVRSIVLAEKRCDEGLTSLRNVSCEDMFKTLHSGALPORGQTOVLCTVTFDSLE 386
Db 304 EHAEVRLITEDKVRPDGRKVDKIRPLDAVVDLPRVHSGGLFTTGQTOALSVLT----- 359

Qy 387 SGIKSDOYVTAINGIKDKNFMILHYEPPYATNIGKTGNNREIIGHALAEKALYVPV 446
Db 360 APMGETOIIDGDDPEYKRFPMHYNFPQYVSVEETERYGAPGRREIGHALAEKALYVP 419
Qy 447 --RDPFTIRVTSVLEBSNGSSSMASACGSLALDMSGVPISSAVAGVAGLVTYTPDK 504
Db 420 SLSEFPVAILRVAEVLBSNGSSSQASICAGTLALMTGVPVIAVPAVAGIAMGLIS----- 473
Qy 505 GEIEDYRLTLGLIEDNGMDFKIAGTNKGTTLQDIDKLPGPPIKIMWALIOQASVA 564
Db 474 -DGNVYTLVTDIQLGLEDHGMDFVAGTRDITLQMDIKIQGTBAELTTEALQAKKA 532
Qy 555 KEIIOINNKTSKPRASRKNENPVETVQVPLSRKAFVGGGYNLKQLQAGTGTISQ 624
Db 533 REIILDVLEATIPERPELPAPAPKIDITIKIDVDKIKIVIGKGETIDKIMETGVKIDI 592
Qy 625 VDEETFSVAPLPYSVMEARDFITEICDDQEQLEFGAVYATYATTEIRDTGVMYLYEN 684
Db 593 DEGNVSIYSSDQAINRAKEIILAGLVBAKVD-----VYRAKVRIEKFAGFVNLF-D 646
Qy 685 MTAVALHNTOLDNERLN 701
Db 647 KTDALVHISEMAWRTN 663

RESULT 14

US-09-738-626-5665
Sequence 5665, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/577484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5665
LENGTH: 753
TYPE: PRF
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5665

Query Match 28.5%; Score 1014.5; DB 10; Length 753;
Best Local Similarity 35.8%; Pred. No. 1.4e-79;
Matches 255; Conservative 125; Mismatches 263; Indels 69; Gaps 19;
Qy 33 DLGNKRLISGGLARFADGSAVVO-SDPTAVMTAVSKTKPSQ-QMPPLVVDYRQRA 90
Db 25 DFGTRTIRFEFGQLRQADGAVTLYLDDDTMLATTASNPREGFDFPLTVDAEERY 84
Qy 91 AAGRIPTYLRREVSTSKELIIRISIRPLFPAGFYDVOYLCMLAVDGVNEPDV 150
Db 85 AAGRIPSGFRREGSPTEALILACRLIDRPLRPTFKGLRNEVOYVTVMSHNPEDYDV 144
Qy 151 LAINGASVALSLSDIPMNGPVAGVARIIGI-----DGEVYVNPTRKENSSTLNLVAGA 204

```

Db 145 VAINGASATREISGLPVSGAVGVMMALVDEKHEGQVAFPTHAQESVFEIIVAGR 204
Qy 205 PKSQ-----IVMLEASA-ENILQO-----DFCAIKVGVYTOOI 238
Db 205 LVERKRGKRTSDVAVMMVEAGASENVVRKVDGAPATEKIVSGLEAAKPFIDILCRA 264
Qy 239 IQGIQOLVETGVTRTPQKLFTF-SPEIVKYTHKLAMERLYAVFTDYEDHKVSHDAVN 297
Db 265 OEGIAQV--GNAKEFP--LFPPTDEVSAVERKVSCKLASLIT--LKAQEDDADYN 318
Qy 298 KIRLDETEQOLKEF-----PEADPVEIIESFNVAEVRSTYLVNKKCDGRDLTSL 350
Db 319 AYMEIEAEELLPKFEASYSASAASK-ERAGYVMMALIVRMLTDFHFDGKGVTDI 377
Qy 351 RVNCEVMPKTLHGSALFORGOTVLTCTVTPDSLESGIKSDQVITAINGIKDKFMFLHY 410
Db 378 RDLAVEVLLIRAHSSSLFERGETOILGVTTIDML-----KMEQQLDSLAPDARKYMHY 433
Qy 411 EPPVATNEIKRVGTGLNRELGHGALAEKALYVPIP--RDPFTIRVTSVLESGSSSM 468
Db 434 NFPPTSTGETRGVSPKREIGHGALAEKALYVPIPSREFFYAIRQVSEALSGNSTSM 493
Qy 469 ASACGSLALMDSGPVSSAVAGVALGVTKDPEKGEI--DYRLTLDTLIGIEDYNGD 525
Db 494 GSVCASTSLYNAGVPLAPVAGIAMGLVS-----GEIDGKTEVVALTDILGAEADAGD 547
Qy 526 MDPKLAGTNKGTALQADIKLPGIPIKIYMEALIOQASVAKKEILOIMNKTISKPRASKE 585
Db 548 MDPKAGADPTITLQDTKLDGIPSKVSDALEBARAKRLITLNTMADVINGPBMK- 606
Qy 586 NGPVETVQVPLSKRAKFGVPGVNLKQLQAEVTGISQVDEETFSVPAPPSVMEHARD 645
Db 607 FAPRTTYKIPAKIGELIGPKGNINALTETGANISIEDDGVFISAADGASAEALIE 666
Qy 646 FITECKDOQOQLEFGAVYTAITTEIRDTGVMTGLYNMTAVLLHNTOLN 697
Db 667 KINALA--NPOLPKVGERFLGTVTKTAFGAFVSLPGRDG-LVHISKGN 714

RESULT 15
US-10-156-761-10060
; Sequence 10060, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10060
; LENGTH: 738
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10060

```

Query Match 28.1%; Score 999.5; DB 15; Length 738;
 Best Local Similarity 34.5%; Pred. No. 2.9e-78;
 Matches 240; Conservative 130; Mismatches 262; Indels 63; Gaps 16;

```

Qy 35 GNRKLEISSGKLARPADSAAVQ--SGDTAVMTAVSKTKSPS--OFMPLVVDYRQKAAA 92
Db 18 GTRTIRFETGLAKOAGSAVAAYLDDDTMVLATTTASKPKPENLDFPLTVDVEBRMYAA 77

```

```

Qy 93 GRIPNTYLRREVGTSDKEIILTSRIIDRSIRPLFPAGYFYDTQVLCNLAAVGVNEPVL 152
Db 78 GKIRGSFRRREGRSBDAILTCRIDRPLRPSFKGLRNEIQVATIMALNPHLYVVA 137
Qy 153 INGSVALSLSDIIPWANGPVAGVIRIGITIDGVYVNPTRKEMSSSTLNLVACAPSQ---- 208
Db 138 INNAASATOLAGLFPSPGIGVVALINGQVAFPTTELEDAVFDVNVVAGRALEDQVA 197
Qy 209 IVMLEASABENILODFCHAIKVGKVTYQOIIQGLVKE---TGVTKRFP--OKLFTPS 263
Db 198 IMVNEARATE-----KTIQLVAGGAEAPTEEVVAAGLDAAKPFIVLCKAQ 243
Qy 264 PEIVKYTHKLAMERLYAVFTDYEDH-----KVSHDEAVNKRLD 302
Db 244 ADLAAKAAKPTGE--FPVFLDYODDVLEALTAIVKSELQALTIAGQDREAELEDRVKEI 301
Qy 303 TEEOLKEFPPEADPYEIIIESFNVAKEVRSIVLNEKRCDCGRDLTSLRVNCEVDMFKT 362
Db 302 AAEKTLRPF--EGREKEISAAVRSITKSLVERVYKDKVRIDRGVTDIRTLAAEVEAI 360
Qy 363 LHGSALFORGOTVLTCTVTPDSLESGIKSDQVITAINGIKDKFMFLHYEPPVATNEIGK 422
Db 361 VHSALFERGSTQLIGVTTNML-----RMEQQLDTLSPVTRKRYMHNFPVSVGETGR 416
Qy 423 VTGLNRELGHGALAEKALYVPIP--RDPFTIRVTSVLESGSSMASACGSLALMD 480
Db 417 VGSFKREIGHGALAEKALYVPIPTREPPYAIRQVSEALSGNSTMGVCASTMGLN 476
Qy 481 SGVPSSAVAGVALGVTKDPEKGEIEDRLTDIIGIEDYNGDMDPKLAGTNKGTAL 540
Db 477 AGVPLKAPVAGIAMGLISQ--EINGETH-YVALTDILGAEADAGDMPKTAGTKEFVTAL 533
Qy 541 QADIKLPGIPIKIYMEALIOQASVAKKEILOIMNKTISKPRASRENGPVETVQVPLSKR 600
Db 534 QDITKLDGIPASVLAALKQDRDKRLHLDVMEADITP--DEMSNAPRTIYKIPVDKI 592
Qy 601 AKFVPGGVNMLKQLQAEVTGISQVDEETFSVPAPPSVMEHARDFITECKDOEOOLE 660
Db 593 GEVIGPKGNINQIQDPTGAETTEDDGTIYGQVGSQABARATINGIANPTMP---E 649
Qy 661 FGAVYTAITTEIRDTGVMTGLYNMTAVLLHNTOL 695
Db 650 VGERYLGTVKTTTFFGAFVSLPGRDG-LHHSOI 683

```

Search completed: January 8, 2004, 10:42:19
 Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2004, 10:33:56 ; Search time 27 Seconds
(without alignments)
2511.071 Million cell updates/sec

Title: US-09-907-907A-42

Perfect score: 3557
Sequence: 1 DGFPLPRDRALTLQVRA.....FAVLHNTQDNERILTLIP 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1365.5	38.4	991	2 T48631	polynucleotide pho
2	1204.5	33.9	703	2 B97123	polynucleotide
3	1190.5	33.5	698	2 B89901	polynucleotide
4	1166.5	32.8	713	2 AE2586	polynucleotide
5	1166.5	32.8	713	2 F97368	polynucleotide
6	1161	32.6	711	2 AD0901	polynucleotide pho
7	1158.5	32.6	749	2 G97782	hypothetical prote
8	1158	32.6	734	2 E91134	polynucleotide pho
9	1158	32.6	734	2 H85979	polynucleotide pho
10	1156	32.5	810	2 G75320	polynucleotide pho
11	1153	32.4	718	2 S74509	polynucleotide
12	1152	32.4	734	2 H65106	polynucleotide
13	1150	32.3	714	2 AC3497	polynucleotide
14	1146	32.2	705	2 AC0424	polynucleotide
15	1145	32.2	704	2 G83950	polynucleotide pho
16	1140.5	32.1	745	2 B71654	polynucleotide
17	1140.5	32.1	775	2 B70320	polynucleotide
18	1138.5	32.0	718	2 AD3355	polynucleotide
19	1137	32.0	719	2 F82831	polynucleotide pho
20	1134	31.9	709	2 B64056	polynucleotide
21	1132	31.8	705	2 S70691	polynucleotide
22	1130	31.8	723	2 AG1603	polynucleotide pho
23	1127	31.7	723	2 AC1241	polynucleotide pho
24	1121	31.5	709	2 B82298	polynucleotide
25	1115	31.3	707	2 B84973	polynucleotide
26	1114.5	31.3	694	2 D86615	polynucleotide
27	1114.5	31.3	694	2 G72009	polynucleotide
28	1112	31.3	707	2 C81161	polynucleotide
29	1111	31.2	706	2 H81943	probable polynucleotide

30	1106	31.1	709	2 S38883	polynucleotide
31	1105	31.1	712	2 B87253	polynucleotide
32	1097	30.8	701	2 G83052	polynucleotide
33	1091	30.7	708	2 A72264	polynucleotide pho
34	1068.5	30.0	695	2 B71463	probable polynucleotide
35	1067.5	30.0	722	2 D70200	probable polynucleotide
36	1066	30.0	702	2 C71269	probable polynucleotide
37	1064	29.9	693	2 G81725	polynucleotide
38	1063.5	29.9	773	2 G86856	hypothetical prote
39	1054	29.6	737	2 D95068	polynucleotide
40	1053	29.6	775	2 D97936	polynucleotide
41	1029.5	28.9	719	2 G81332	polynucleotide
42	1016.5	28.6	897	2 T06540	polynucleotide
43	1005.5	28.3	739	2 T10932	polynucleotide
44	1000.5	28.1	777	2 T44900	probable polynucleotide
45	998.5	28.1	773	2 H87015	hypothetical prote

ALIGNMENTS

RESULT 1				
T48631 polynucleotide phosphorylase - Arabidopsis thaliana				
N:Alternate names: protein T15N1.70				
C:Species: Arabidopsis thaliana (mouse-ear cress)				
C.Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000				
C.Accession: T48631				
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le				
submitted to the Protein Sequence Database, April 2000				
A:Reference number: 224493				
A:Accession: T48631				
A>Status: preliminary				
A:Molecule type: DNA				
A:Residues: 1-991 <BEV>				
A:Cross-references: EMBL:AL163792				
A:Experimental source: cultivar Columbia; BAC clone T15N1				
C:Genetics:				
A:Map position: 5				
A:introns: 107/3; 159/3; 216/3; 258/3; 312/3; 352/3; 398/3; 453/1; 495/1; 538/3; 612/1;				
A>Note: T15N1.70				
Query Match				
Best local similarity 38.4%; Score 1365.5; DB 2; Length 991;				
Matches 287; Conservative 138; Mismatches 231; Indels 33; Gaps 11;				
QY	24	SAGSRAV----	AVDLGNRKLEISGKLARPADGSAVQSDTAVMTAVSKTPSPSQF	78
DB	44	SAGTKILSFKEBFEVSGRVVS	FETGKIFARFANGSVVLGMDETKVLSIVTCAKTDSPDF	103
QY	79	MLVVDYRQKAAAGRIPTNTY	LRREVGTSDKEILTSRIIDSIRPLFPAGFYDTQVLCN	138
DB	104	LPLTVDOEKQYAOGLFNTY	MRREGAPKEREELCGRIIDRPIPLPFTGVEHQIIVAS	163
QY	139	LLAVGVNRPVLAINGASVAL	SLSDIPMNPGVAVRGIIIDGVEVNAVPTKREMSSTLN	198
DB	164	VLSDDGKDPDIPILANAS	SALMLSDVWGPGLVIRIGICGQVNPVMDLSSDLN	223
QY	199	LVAAGAPKSQIVMLEAS	ENITLQDPCHAIKY---GVKTYQOIIQGIQOLVKETGVYKR	254
DB	224	LIYA-CTRDKTMWIDVGS	REISBKDLAALRLAPBAVKYLDPOI---RLAKRAGKQKK	278
QY	255	TPQKLFPSPEIVKTKL	AMERLYAVFTDYEHDKVSDAENKIRLDTESQLEKPEA	314
DB	279	EYKLSMLSDTKLRK	VDLAATRIESVPTDPSYGFEGEALDNIQKQVRYFEEBQOE	337
QY	315	DPEYIESFNVAKEPS	IVINAEYKRCDDGLTSLRVNCEVDMFKTLHGSALFORQOT	374
DB	338	SLSLIPRAVDYVRK	KVARSMTISDGFVDSGHVDEVRPIYESHYLPALHSAALFSGDT	397
QY	375	QVLCTVTFDSLS	ESGISKSDQVITAINGIKDKNFMHYEPPYATNIEIGKVTGALNRELGHG	434
DB	398	QVLCTVTLGAD	PAEAQSDLSIV---GPKKKFMHYSPFYCTNBEVGRGGLNREVEVGHG	453

435 ALAEKALVPIPRD--PFTIRVTSVLENGSSSSMASACGSLALMDSGVPISSAVAGV 492
454 TLAEKALLAVLPPEAPPTIRINSEVMSSDGTSMSVCGSALMDAGIPLAAHAGV 513
493 AIGLVTKDPEKSEIEDYRLITLIGIEDYNGMDPKIAGTNKIGITALQADIKLPGIPIK 552
514 SVGLITVDPSGSGIKQYRIYVTDILIGLEDHGDMDPKIAGTRDGTALQDLIDKAGIPLD 573
553 IYMAIIOQASVAKKEIILQIMNKITSKPRASKENGPVETVQVPLSKRAKRVGSGYNLK 612
574 IVCSLENAERARLQIIDLHERNINSPRGQAVSPRLATLKYNSDLRTILIGPWGLKR 633
613 KLAETGVTSIQVDEEFPSPAPTPSVHMBAR---DFTIEICKDQEQOLEFGAVYTATI 669
634 KIEVETGARLS-IDNGITITVAKNQDVMKQOEVDPII-----GRELVGGVYKQTV 685
670 TEIRDYGVMTCLYPMNTAVLLANTQDNE 698
686 SSIKEYGAFVE-FPGCGQGLHMSLSHR 713

RESULT 2

polyribonucleotide nucleotidyltransferase [Imported] - Clostridium acetobutylicum
C/Species: Clostridium acetobutylicum
C/Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C/Accession: B97123
R/Nolling, U.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A/Reference number: A96900; MUID:21359325; PMID:21359325
A/Accession: B97123
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-703 <KUR>
A/Cross-references: GB:AE001437; PIDN:AAK79773.1; PID:G15024781; GSPDB:GN00168
A/Experimental source: Clostridium acetobutylicum ATCC824
C/Genetics:
A/Gene: CAC1808
C/Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

Query Match 33.9%; Score 1204.5; DB 2; Length 703;
Best Local Similarity 39.6%; Pred. No. 4-5e-68;
Matches 266; Conservative 125; Mismatches 258; Indels 23; Gaps 10;
33 DLGNRLKLEISSGKLARFADGSAVVOGDTAVMTAVSKTKPSPS-QEMPLVVDYRQKAA 91
8 DIAGRKLVCECGKGMNSCAMFISYGDYVMMVNNVNASEKREGIDFPLSIEYERQYS 67
92 AGRIPTVTLREVGTSDEKEILTSRIIDRSIRPLPAGFYDTQVLCNLLAVDGNNEPVYL 151
68 VGKIPGCVKKEGSEKSIHARAIIDRLPLPFGKGRNDVQVCTWASVEONLPEIL 127
152 AINGASVALSLSDIPMNGPVGAVRIGTIDGXYVNPTRKEMSSSTLNLVAGAPSOIVM 211
128 AMNGASVALSLSDIPFTTPVAVSVGCTDGKFLVLPFTLEBEKSSLDLTVCAITNERMM 186
212 LEASAEENILQODFCHAIKVGKVTQOIIQGIQOLVETGVTKTPQKLFPSPEIYKXTH 271
187 LEAGADEIPEDLMTAIDIFGNACQDIVAFQEKAMKEGKEMVPE-LYHFKELIKQVT 245
272 KLANERLYAVFTDYEHDKVSDAEVANKIRLDTBEQLEKKEPBADEIETIESPNNVAKGV 331
246 EFAPESIEKIM--ITTDDEBNRLRLREIKESINEFAKYPD-DGADIDEVAVYTLQKKV 302
332 RSVTANEXKCDGRDLTSLRVNSCEVDMFKTLHGSALEFORQOTVLCVTFPDSLESGTIS 391
303 RNMLIKERRPDGRFDEIRIPISCDVLLPRTSGGLFTRGLTQVMYTTIL---GPIGD 358
392 DQVITAINGIQDKQFMILYEPFPYATNIEIGKVTGLNREELHGALAEKALVPIPR-RDF 449

Db 359 AOVIDGLVEBSKRYMHHNFPYSTGEVYKPLGNRRREIGHALAEKALVPLIPSEEF 418
450 PFTIRVTSVLENGSSSSMASACGSLALMDSGVPISSAVAGVIGLVTKDPEKSEIED 509
419 PFTIRVTSVLENGSSSSMASACGSLALMDSGVPISSAVAGVIGLVTKDPEKSEIED 476
510 YRLITDILIGIEDYNGMDPKIAGTNKIGITALQADIKLPGIPIKIMBAIOQASVAKKEIL 569
477 ---ITDILIGLEDHGDMDPKIAGTRDGTALQDLIDKAGIPLD 533
570 QIMNKITSKPRASKENGPVETVQVPLSKRAKRVGSGYNLKQLAETGVTSIQVDEEFT 629
534 EKVVACINBRKELSTYAPRAYINIDTDIRLIGTGKTKINKIETGVKIDIRDDGT 593
630 FSVAPTPSVHMBARDFITEICKDQEQOLEFGAVYTATITBTDGVMYKLYPMNTAVL 689
594 VFVLSSDASANRLKMIIDTLTD---YKVGVEVYLGKVTKITNFGAFVPLDCKEG-L 647
690 LHNVTQDNERLN 701
648 WHISKDINKYN 659

RESULT 3

polyribonucleotide nucleotidyltransferase [Imported] - Staphylococcus aureus (strain N31)
C/Species: Staphylococcus aureus
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C/Accession: B89901
R/Kirouda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc, ma, A.; Mutsaers, H.; Y. Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; i C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A/Reference number: A89758; MUID:21311952; PMID:11418146
A/Accession: B89901
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-698 <KUR>
A/Cross-references: GB:BA000018; PID:G13701074; PIDN:BAH42369.1; GSPDB:GN00149
A/Experimental source: strain N31
C/Genetics:
A/Gene: pnpA
C/Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

Query Match 33.5%; Score 1190.5; DB 2; Length 698;
Best Local Similarity 39.1%; Pred. No. 3.4e-67;
Matches 273; Conservative 118; Mismatches 249; Indels 59; Gaps 12;
24 SAGSRAVAVDLGNRLKLEISSGKLARFADGSAVVOGDTAVMTAVSKTKPSPSQFMPLVV 83
2 SQEKVTKTEMAGRSULTIBTGQAKONGAVLVRYGTVVLSRTAKKERPDGFPPLTV 61
84 DYRQKAAAGRIPTVTLREVGTSDEKEILTSRIIDRSIRPLPAGFYDTQVLCNLLAVD 143
62 NYBEKVVAAAGRIPOGFKRGRPDATLTARLIDRIRPLPFGYGVHDVQIMMVLASD 121
144 GVNBPDLANGASVALSLSDIPMNGPVGAVRIGTIDGXYVNPTRKEMSSSTLNLVAG 203
122 PDGSPQAAAMIGSSMALSVSDIPQGIAGVNGYIDGKYYINPTVEKEVSRLDLVAG 181
204 APKSOIYMLEASAEENILQODFCHAIKVG-----VKYTOQIIQGIQOLVETGVTKTP 256
182 -HDAVAVNVAAGASIEIQEMLLEIFGHEIQGLVFOQDIYHIDPVQOE----- 232
257 QKLFPS-----PIVYTKHLAMERLYAVFTDYEHDKVSDAEVANKIRLDTBEQLEK 309
233 ---FIPARDEALVERVKSITBEKGLKETYLT-----DKQORDENDNLK---EETVNE 281
310 KFPDADP-----YEIESPNNVAKGVPRSVLANEYKCGRDLTSLRVNSCEVDMFKTL 363
282 FIDEDDENELLKEVYALINELVEVRRLADEKIRPDGRKDEIRLDSSEVGLPRT 341

241 LENEMGLAT-ELTAYKITEQARYAANDAVKTKYKAHFLPBEGBKAYSPEIGIVFX 299

QY 268 KYTHLAMERLVAFPTDYE-HDKVSRDEAVNKIRLDTEEQLEKFPPEA--DPYEIIIESFN 324

```

Db      241  LBNEMLGLAET-ELRTAVYKTEKARVAAVDAVKTAKVAHFLPBEGBEAKXSPBEIGAVK 299
Qy      325  VVAKVEFRSIVLNTXKRCDCGRDLTSLRNVSCEVDMFKTLGSAALFORQOTVLCTVIEDS 384
Db      300  HLQAKIVAMNVLDTYSRIDGRDLSTVRPIVSEVGLIPFTHSALPRTRETAIVATIGT 359
Qy      385  LESGIKSDQVITAINGIKDKNFMLYHEPPYATNIEIGVTGLANREILGALAEKALYV 444
Db      360  GE----DEQYVDSLGMWKERFLHYNPYPYSVBETGMSGPCGRREIGHGLAMRALRPM 415
Qy      445  IP--RDPFFTIRVTSVLESNSSSMASACGSLALMDSVPISSAVAGVALGLVTTDP 502
Db      416  LPTAEQFPYTLRVVSEITESSGSSSMATVCTSLALMDAGVPLAKPVAGIAMGIL---- 471
Qy      503  EKGIEEDYRLTLDIGIEDYDMDFKIAGTNKGTITLQADIKLPGIPYIKLYMEALIQAS 562
Db      472  ---EGERPAVLSDIGDEDDHJGDMDFKAVAGTADGITSLOMDIKIAGITIEEIMKIALBOAQ 528
Qy      563  VAKKEIIOINMKITSKPRASRKENGPNVETVQVPLSKRAKVVGPGNYLKKLOAFETGTI 622
Db      529  GGRKRIIIEEMNANALITESRGOLGEFAPRIEIVNNIPDKIREVIIGSGKRIEIVEKTGAKI 588
Qy      623  SQVDEETSVAPFTSVHHEARDFTIEICXDDQEOQLFEFNAVYATITIEINDTGMYKLY 682
Db      589  NIEBDGYIKYIASASKEIEAARKWIHISIVAP-----EVGOIYEGTIVKTDPGAFAVNF 643
Qy      683  PNMTAVLLHNTQLODNERL 700
Db      644  GARDG-LVHISQULASERV 660

```

RESULT 6
AD0901

polynucleotide phosphorylase [imported] - *Salmonella enterica* subsp. *enterica* serovar Typh
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typh
A:Note: This species has also been called *Salmonella typhi*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AD0901
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
C.H.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
S.; Mole, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quall, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AD0901
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-711 <PAR>
A:Cross-References: GB:AL513382, PIDN:CAD07802.1, PID:G16504350, GSPDB:GN00176
C:Genetics:
A:Gene: STY3463
C:Superfamily: polynucleotide nucleotidyltransferase alpha chain

Query Match	32.6%	Score 1161,	DB 2;	Length 711;
Best Local Similarity	39.9%	Pred. No.2.6e-65;		
Matches 270;	Conservative 118;	Mismatches 245;	Indels 40;	Gaps 11;

Polynuclear Aromatic Hydrocarbons (PAHs) are a class of organic compounds consisting of multiple fused benzene rings. They are commonly found in the environment, particularly in the context of air pollution and environmental health. PAHs are known for their potential carcinogenicity and are often studied in relation to their sources, distribution, and effects on human health.

A:Note: this species has also been called *Salmonella typhi*.
 CDate: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 CAccession: AD0901
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
 ch, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
 , S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quall, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
 A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AD0901
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-711 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD07802.1; PID:G16504350; GSPDB:GN00176
 C:Genetics:
 A:Genes: STY3463
 C:Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

Query Match 32.6%; Score 1161; DB 2; Length 711;
 Best Local Similarity 39.9%; Pred. No. 2,6e-65;
 Matches 270; Conservative 118; Mismatches 249; Indels 40; Gaps 11;

QY	35	GNRKLEISGKLAPADGSAYVQSGDTAVMTAVNSKTRPSQ-FMPLVVDYRQAAAG	93
DB	12	GQHTVLTETGMARQATAAVWVMSMDTAVFVTVVQKAKKAGQDFPLTVVQERTVAAG	71
QY	94	RIPTYLREWGTSGKEILTSRIIDRSIRLPFPAGYFDTQVLCVLAVDGNBPDVAT	153
DB	72	RIPGFFPRRGRPSGETLITARIIDRPVAPLEPESFVNEVOVIATVAVSNQVNPDIYAM	131
QY	154	NGASVALSLSDIPMNGPVGAIVAGIIDEYVNPFRKEMSSSTLNLVYAGAPKQIYWL	213
DB	132	IGASALSLSGIFPGPIGAARVGIYNDQYVLPFODELKESKDLVVAAGT-AAVAVLVE	190
QY	214	ASAEYIILQDPFHAKVGVKTYQQLIQGIQQLVETGYTKRTPQLFTFSPSEIYK-----	268
DB	191	SEAEILSDTMIGAVFGEQGVVIAINDLVKAGKPRMDWQ-----DEAVNDALNA	244

```

OY      269 YTHKLAAMEL---YAVFTYDEHDKVSRDPAVVKJRLDTEBOLKEKFPDAPDYEIIESFNV 325
Db      245 RVALLESRSIDAIRI-----TDKQERYAQDVILKSEITIEOLIADETLDANIELGILHA 299
OY      326 VAKEVERSIVLNEYKRCGGRDLTSLDRNYSCEVDEKTLHGSALFORQOTVLTCTPDSL 368
Db      300 IENGVVRSLVAGEPRIDREKMDIRGLDVRGVLPRTHSALPTRGTOVLATLTGT- 358
OY      386 ESQIKSDQVITAINGIKDKNFMLAHPPEPPYATNEIGKTYGLNREIEIGHLAELALYPVI 445
Db      359 ---ARAQVLDLBMGRRTDSFLFHNPPYSVSGEIVMGVSPKREIEIGHGRALKGVLAIVM 415
OY      446 P--RDEPPTIRVTSERTLESNGSSMASACGSLALMDSGVSISSAVGVALIGLTTKTPDE 503
Db      416 PDMDKFPYTRVVSSETTESNGSSMASVCGASLALMDGVPIKRAAVGAMIGLVKEED-- 473
OY      504 KGEIEDYRLTLTLGIEDYNGMDPKIAGNTKGITALQADIKLPGLPIKTYMEALIQASV 563
Db      474 ----NYVVLSDILGDEBHLGMDMDKVKVSGRSIGLSALQMDIKIGITKEIMVQLNDAKG 528
OY      564 AKKEIILQINNKTKISKRPARSKENGGVENVVOVPLSKARFVGPBGVYLKKLQAEVTYIS 623
Db      529 ARHLITGVMEQALNABRGDISFAPRHTTKISTDKIKDVIYKGGKGSVIRALTTESTGTTIE 588
OY      624 QVDEETFSVAPTPSVMEHARDPTEIKCDQBOQLEFGAVYATLTTEIRDTGVVVKLYP 683
Db      589 IEDDGVTKLAATGEKAKAYAIRIEI-----TAIEVGRITNSKVTRIYDFGAFVAIGG 643
OY      684 NMTAVVLLHNTQDLNERTL 700
Db      644 GKEG-LVHISQILADKKV 659

```

RESULT 7

G97782
hypothetical protein pnp [imported] - Rickettsia conorii (strain Malish 7)
C|Species: Rickettsia conorii
C|Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C|Accession: G97782
R|Ogata, H.; Audic, S.; Reneato-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro-
Science 293, 2093-2098, 2001
A|Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A|Reference numbers: A97700; MUID:21442074; PMID:11557893
A|Accession: G97782
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-749 <KUR>
A|Cross-references: GB:AE006914; PIDN:AAL03201.1; PID:g15619751; GSPDB:GN00173
C|Genetics:
A|Gene: pnp
C|Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

**Best U
Matche**

[illegible]

QY 270 THLD---AMERLYAVFTVDEHDKVSRDEAVNKIRLDTBEOLKEKPEADPEYIEIESFNVA 326
DB 245 IEDLVFKYIEKQAFKISKOR--STNLDLPEKVLTHVSDIENK--KSYNYQIESALKAI 301
QY 327 AKRPSIVLNEYKRCODGRDLTSLRNVSCEVDMFKTLHGSALFORQOTVLTCTVTF--DGL 385
DB 302 ESDILNBELEKRRIDGNTSTDIRQACIGLPSHSGALFTRETSOLSVSTTGTSL 361
QY 386 ESGIKSDOYITAINGIDKKNFMLEPFPYATNEIGVTGLNRELGHALAEKALPYI 445
DB 362 D-----EQVDSLEGGYKRFMLNTYFPFYSVNEAMPKAPSRREYGHGLARALNPLI 416
QY 446 PR--DPPFTIRVTSEVLNNGSSSMASACGSLALMDSGVPISSAVALGIVTKTDE 503
DB 417 PKNVQPPYSIRVAAETTESNGSSSMATVCCSIALWAGVPIKAPVAGIAMGLVK----- 471
QY 504 KGEIEDYRLTLTDLGIEDYNGDMDFKTAGNKITLQADIKLPGIPKIVMEALIQASVAK 563
DB 472 --BGRKPAVALSDILGSDYFGDMDFVAGTSBGITLQMDIKISGVDFIMKVALAEQARL 529
QY 564 AKKEILQIMNKTISKPRASKENGPPVETVQVPLSKRAKFPVPGGNLKLQATGVTISV 623
DB 530 GRHLIEQNNKIVISKNNELSKAPSTTTIKIDKIDRIIGPGKIIKEICTSGAKID 589
QY 624 QVDEETSVFAPTPSVMEHARDFTIEICKDOQEOLEFGAVYTATITEIRDTGVWVKLP 683
DB 590 ISPDGTVSVYASDRDLKVALDKIKAIIVEP-----EIGEIFNGTVVVKLDSGAFINYG 644
QY 684 NMTAVLHNTQDLNERL 700
DB 645 NKDG-FVHISEVGERI 660

RESULT 8
B91134

polynucleotide phosphorylase [imported] - Escherichia coli (strain O157:H7, substrain R1)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: E91134
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gsawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E91134
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-734 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA837468.1; PID:g13363518; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: BCG4045
C:Superfamily: polynucleotide nucleotidyltransferase alpha chain

Query Match 32.6%; Score 1158; DB 2; Length 734;
Best Local Similarity 40.0%; Pred. No. 4.2e-65;
Matches 270; Conservative 117; Mismatches 252; Indels 36; Gaps 11;

QY 35 GNRKLEISGKLARFADGSAVVGSDTAVMTAVSKTPSPSQ-FMPLVVDYRQKAAAG 93
DB 35 GQHTVLTETGMARQATAVAVVSMDDTAVFVTVGQKAKPGQDFPLTVNQEERTYAG 94
QY 94 RIPTNVLREVGSDEKILTSRIIDRSIRPLPAGFYDYOVLNCLLAVDGNBPDVLA 153
DB 95 RIFGSEFRBGRPSBESITLARLIDRPIPLPFBGFAVNAVQVATVAVSNPQVNPVIAV 154
QY 154 NGASVALSLSDIPMNGPVGARIGIIDGVVNPTRKEMSSSTLNLVAGAPKQIWMLE 213
DB 155 IGASVALSLSGIPFNGPIGAARVGINDOYVNPQDELKESLTLVAGT--EAAVLWME 213
QY 214 ASANENLOODFCHAIKVGKVTQOIIQGIQOLVKETGVTKTPQKLTFTSP---EIVYKT 270
DB 214 SEAEHLISBDMGAVVFGHEQOQVVIQINELVKEAG---KPRMDQCEPVMEALNARV 269

QY 271 HKLAMERL---YAVFTVDEHDKVSRDEAVNKIRLDTBEOLKEKPEADPEYIEIESFNVA 327
DB 270 AALAEARLSDAARI-----TDKQRYAQVDVAKSETIATLAEDELDTBNEIGELIAHE 324
QY 328 KEVPSIVLNEYKRCODGRDLTSLRNVSCEVDMFKTLHGSALFORQOTVLTCTVTF--DGL 385
DB 302 ESDILNBELEKRRIDGNTSTDIRQACIGLPSHSGALFTRETSOLSVSTTGTSL 361
QY 386 ESGIKSDOYITAINGIDKKNFMLEPFPYATNEIGVTGLNRELGHALAEKALPYI 445
DB 362 D-----EQVDSLEGGYKRFMLNTYFPFYSVNEAMPKAPSRREYGHGLARALNPLI 416
QY 446 PR--DPPFTIRVTSEVLNNGSSSMASACGSLALMDSGVPISSAVALGIVTKTDE 503
DB 417 PKNVQPPYSIRVAAETTESNGSSSMATVCCSIALWAGVPIKAPVAGIAMGLVK----- 471
QY 504 KGEIEDYRLTLTDLGIEDYNGDMDFKTAGNKITLQADIKLPGIPKIVMEALIQASVAK 563
DB 472 --BGRKPAVALSDILGSDYFGDMDFVAGTSBGITLQMDIKISGVDFIMKVALAEQARL 529
QY 564 AKKEILQIMNKTISKPRASKENGPPVETVQVPLSKRAKFPVPGGNLKLQATGVTISV 623
DB 530 GRHLIEQNNKIVISKNNELSKAPSTTTIKIDKIDRIIGPGKIIKEICTSGAKID 589
QY 624 QVDEETSVFAPTPSVMEHARDFTIEICKDOQEOLEFGAVYTATITEIRDTGVWVKLP 683
DB 590 ISPDGTVSVYASDRDLKVALDKIKAIIVEP-----EIGEIFNGTVVVKLDSGAFINYG 644
QY 684 NMTAVLHNTQDLNERL 700
DB 645 NKDG-FVHISEVGERI 660

RESULT 9
H85979

polynucleotide phosphorylase [imported] - Escherichia coli (strain O157:H7, substrain ED)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 27-Nov-2001
C:Accession: H85979
R:Berria, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamocasts, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85979
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-734 <STO>
A:Cross-references: GB:AE005174; NID:g12517775; PIDN:AAG58930.1; GSPDB:GN00145; UNGP:245
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: pnp
C:Superfamily: polynucleotide nucleotidyltransferase alpha chain

Query Match 32.6%; Score 1158; DB 2; Length 734;
Best Local Similarity 40.0%; Pred. No. 4.2e-65;
Matches 270; Conservative 117; Mismatches 252; Indels 36; Gaps 11;

QY 35 GNRKLEISGKLARFADGSAVVGSDTAVMTAVSKTPSPSQ-FMPLVVDYRQKAAAG 93
DB 35 GQHTVLTETGMARQATAVAVVSMDDTAVFVTVGQKAKPGQDFPLTVNQEERTYAG 94
QY 94 RIPTNVLREVGSDEKILTSRIIDRSIRPLPAGFYDYOVLNCLLAVDGNBPDVLA 153
DB 95 RIFGSEFRBGRPSBESITLARLIDRPIPLPFBGFAVNAVQVATVAVSNPQVNPVIAV 154
QY 154 NGASVALSLSDIPMNGPVGARIGIIDGVVNPTRKEMSSSTLNLVAGAPKQIWMLE 213
DB 155 IGASVALSLSGIPFNGPIGAARVGINDOYVNPQDELKESLTLVAGT--EAAVLWME 213
QY 214 ASANENLOODFCHAIKVGKVTQOIIQGIQOLVKETGVTKTPQKLTFTSP---EIVYKT 270
DB 214 SEAEHLISBDMGAVVFGHEQOQVVIQINELVKEAG---KPRMDQCEPVMEALNARV 269

Db 214 SEABLLSDQMAGVAFGEHQOQVVIQINELVKEAG-----KPRWDQPEPVEBNALNARV 269
 Qy 271 HKLAMERL---YAVFTDYEHDKVSRDEAVNKRILDTERQLEKKEPPEADPYEIIISFNVA 327
 Db 270 AALMBARLSDVYRI-----TDKQERYAQVDYIKSETIATLAEDETTDENELGELIHAIE 324
 Qy 328 KEVERSIYLVNEYKRCDSGLTSLRNVSCEVDMFKTLGASALFORQOTVLCVTVPDSIES 387
 Db 325 KAVVRSRYLAGEBPRIDREKQIMIRGLDVRTGLPRTGHSALFTRGETQALVTATIGT--- 381
 Qy 388 GIKSDQVTAINGIKDKKFMILHYEPFPYATNEIGKVTGLANRELGHGLAKALYLPVP- 446
 Db 382 -ARAQVDELMTGERTDTFLFPHYFPYVSGETGVGSPKREIIGHGLARGLVLAWPD 440
 Qy 447 -RDPFTRVTSVLESGSSSMASACGSLALMDSGPVSSAVAGVAGLVTKTPDEKG 505
 Db 441 MDKPPYRVVSVETIESGSSSMASVCGASLALMDAGPITKAAVAGLNGLVKEGD----- 496
 Qy 506 EIEDYRLTLIGIEDYNGDMDFKIAGTNKGITLALQADIKLPGIPIKIWEALQOASVAK 565
 Db 497 ---HYVVLSDILGDEBDHLDGMDPKVAGSRDGSIALQMDIKIEGITKEIMQVALLAKGAR 553
 Qy 566 KEIIQINAKTISKPRASKEKNGPVETVQVPLSKRAKVFPGGINKKLQAEFTVTSQV 625
 Db 554 LHIIGVMEQAINAPRGDISSEFAPRIHIKINPDKIKDVIKGGSVIRALTEETGTTIEIE 613
 Qy 626 DEETFSVAPPTPSVMEHARDPITEICKDQEOQLFEFVATATITEIRDTGVMTLYPNM 685
 Db 614 DDGTAKLAATDGEKAKHARIIEI-----TAEIEGVVYGVKTRIVDFGAFVAGIGGK 668
 Qy 686 TAVLHNTQDNERL 700
 Db 669 EG-LVHISQIADKV 682

RESULT 10

675320
 polynucleotide phosphorylase - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: G75320
 R:White, O.; Eksen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 Science 286, 1571-1577, 1999
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: G75320
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-810 <WHI>
 A:Cross-references: GB:AE002042; GB:AE000513; NID:g6455848; PIDN:AAPI1608.1; PID:g645985
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR2063
 A:Map position: 1
 C:Superfamily: polynucleotide nucleotidyltransferase alpha chain

Query Match 32.5%; Score 1156; DB 2; Length 810;
 Best local similarity 38.1%; Pred. No. 6.5e-65;
 Matches 261; Conservative 138; Mismatches 236; Indels 50; Gaps 14;

Qy 34 LGNRKLEISSGKLARFADGSAVVOGSDTAWVTVASVKTKPSQFMPVVDYRQKAAAG 93
 Db 41 LGGRELSTETKLAALVSGSVTVRGDTLLVTAQASDTQSKDPLPLTFVEERHVAVG 100
 Qy 94 RLPNRYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFDTOVLGNLLAVDGVNPDVLA 153
 Db 101 KIPSGFORREGRPEKALISARITDROIRPLFPKGRYETVITVVSADQONAPDVIGP 160
 Qy 154 NGASVALSLSDIPWNGPVAVRIGIIDEVYVNPTRKEMSSSTNLVVAAGPKSQIVMLE 213
 Db 161 IGAALALISIDIPWAGPACVAVGQIDQGVVNPFTTEQLTRSRMDLVVAGT-REAVMVE 219

Qy 214 ASAENILQDFCHAIKGVKVTQOIIQGIOLVKEIS-----VTKRTPOKLFTPSPEIV 267
 Db 220 CGAQTVSEDDLVGALIEFAHMEQVILIEQMAEVEHEKFNPLAEGBANDYV----- 274
 Qy 268 KYTHKLMERYL-AVFTDYHDKVSRDEAVNKRIL-----DTEQLKEKPEADPYEIIIE 321
 Db 275 ELTEKAKAAGLRDALLRHGKKDSARLKLKRLNGILIEGVVDPDAEGSAELTQA-----LKD 330
 Qy 322 SPNVAAKEVRSIYLVNEYKRCDSGLTSLRNVSCEVDMFKTLGASALFORQOTVLCVT 381
 Db 331 AFGKVERREIRRLLEENLADGDSKTVRPIWEARPLPALGSAVFTGETVGLVTT 390
 Qy 382 FDSLESIGKSDQV-----ITANGIKDKNFMILHYEPFPYATNEIGKVTGLANRELGHGALA 437
 Db 391 L-----GTERDEILIDULTAESGD---FLHNFPFYSIGVYKRMGOSRREIGHNLA 442
 Qy 438 EKALYLPVP--RDPFTRVTSVLESGSSSMASACGSLALMDSGPVSSAVAGVAG 495
 Db 443 KRAIRAVLPSEEPFVYIRVVGDLVESGSSSMGTVCAGTSLMDAGVPLKAPVAGVAMG 502
 Qy 496 LVTKTPEKEIEDYRLTLIGIEDYNGDMDFKIAGTNKGITLALQADIKLPGIPITVM 555
 Db 503 LVMEGD-----NYRVLTLIGLEBDLGDMDPKVCGTAAGVITALQMDIVGGLTPQIMR 555
 Qy 556 EALQOASVAKKEIIQINAKTISKPRASKEKNGPVETVQVPLSKRAKVFPGGINKKLQ 615
 Db 556 EALAQAKEGRHLHIGKAEVLAPRAELSPATHILSKINPELIGVITPGKQVARELE 615
 Qy 616 AETGVTSQVDEETFSVAPPTPSVMEHARDPITEICKDQEOQLFEFVATATITEIRDT 675
 Db 616 A-MGAQVTEBEDGTVRIFSASGESAEVAKARIEAVTK-----BAKVEEFGFTVAKIAPF 669
 Qy 676 GVAWVLYPNMTAVLHNTQDNERL 700
 Db 670 GAFVNLFPQDG-MLHISQISEQRV 693

RESULT 11

574509
 polynucleotide nucleotidyltransferase (EC 2.7.7.8) alpha chain - Synechocystis sp. (6
 N/Alternate names: protein sll1043
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
 C:Accession: S74509
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O, K.; Okumura, S.; Shino, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 sp.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S74509
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-718 <KAN>
 A:Cross-references: EMBL:D08089; GB:AB001339; NID:g1651650; PIDN:BAAL6661.1; PID:g165173;
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Gene: pnp
 C:Superfamily: polynucleotide nucleotidyltransferase alpha chain
 C:Keywords: nucleotidyltransferase

Query Match 32.4%; Score 1153; DB 2; Length 718;
 Best local similarity 37.5%; Pred. No. 8.3e-65;
 Matches 257; Conservative 131; Mismatches 261; Indels 36; Gaps 10;

Qy 31 AYDLGNRKLISGKLARFADGSAVVOGSDTAWVTVASVKTKPSQFMPVVDYRQKAA 90
 Db 7 SISFDGDIRLKNGTILAPQAGSVLQSGDTAVLVTRAKGADGIDFLPLTYDYERLY 66
 Qy 91 AAGRIPNRYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFDTOVLGNLLAVDGVNPDV 150

Db 67 AAGRIPGFGLRBRGPRPEKATLISRLIDRLPLPFPFHLMDLQIVATTLSMDEEVPDV 126
 Qy 151 LAINGASVALSLSDIPFPMNGPVAGVIGIDGEEVYVPTKEMSSSTLNVAGAPKSOIV 210
 Db 127 LAVTASAVAVILAOIPFKGPMMAVAVGVLDGDDFIITFTRVHNGDLVLVAGTP AGIV 185
 Qy 211 MLEASAEMLIQQDFCHAIKVGKYYTQOIIOGIQOLVKETGKTRPPQKLFTRSP--EIV 267
 Db 186 MVEAGANQIPQDITIEAIDFGYEAQVQDLINAGRELMTDGLTATS-----PPVNTAV 241
 Qy 266 KTHGLAMERLYAVFDYHDYKVSDEAVNKRILDTBEOLKEKFPADYE----- 318
 Db 242 EPIANRASKKITTIVGQFDLGKGRPALDEKATEVEYALAELEPTDPVKQSVDEPQL 301
 Qy 319 IIESFNVAKEVRSIVLNEYKRCQGRDGLTSLRNVSCEYDMF-KTLHGSALEFORGOTVL 377
 Db 302 VGNLYKALTKLMRKQIYVDGVRVDRKLEQVAPRPSCEGFLPRKRVHSGGLFRNGLTQVL 361
 Qy 378 CTVPDLSLSGKISQDVITAINGIKDKNFMHYEPFPVATNEIGKVTGLNREELGHGALA 437
 Db 362 SLAT---LGSQDAQADLADLHPDEKRYLHHYNFPYVSGARPMRSPGRREIGHGALA 418
 Qy 438 EKALPVPVIR--RDPFTTIVTSEVLESNSSSMASACGSLALMDSGVPISSAVAGVAG 495
 Db 419 EKALIPVLPPOBFPVYVAVSEVLSNGSTSGVSGSTLALMDAGVPIKKPVSGAMG 478
 Qy 496 LVTKTDPKGEIEDYRLLDIGIEDYNGDMDFKTAGTKGITALQADIKLPGPIKIVM 555
 Db 479 LKEGD-----EIRILIDIOGIEDFLGDMDFKAGTDSGITLALQMDKIKGLSMEVVS 531
 Qy 556 EAIQOASVAKKELIQMNTKISPRASRKNRPNVETVQPLSKAFVPGGYMLKKLO 615
 Db 532 KAIMQALPARHLIDLMKLTITREPRELSPAPRLITLKEIPEHIGMTVPGGKTIKGT 591
 Qy 616 AETGTTISQVDEETSVFAPTPSVWHEARDFITEICKDQEQOLBEGAVYTTATTEIRDT 675
 Db 592 EOTSCIKIDADGTYVIASSBGRARERARQMTYMTNR-----KLNEGVEYLLGRVRIIPI 646
 Qy 676 GVMVLYPMNTAVLHNTQLDNERL 700
 Db 647 GAFVEVLPGEKG-MHISQLTGGRV 670
 RESULT 12
 H65106
 polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) alpha chain - Escherichia coli (E
 N) alternate names: polyribonucleotide phosphorylase
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence revision 23-Jan-1998 #text_change 01-Mar-2002
 C:Accession: H65106; A26118; B23984
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: H65106
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-734 <BLAT>
 A:Cross-references: GB:A0000397; GB:U00096; NID:G2367199; PIDN:AC76198.1; PID:G1789555;
 A:Experimental source: strain K-12, substrain MG1655
 R:Regnier, P.; Grunberg-Manago, M.; Portier, C.
 J. Biol. Chem. 262, 63-68, 1987
 A:Title: Nucleotide sequence of the pnp gene of Escherichia coli encoding polynucleotide
 in S1.
 A:Reference number: A26118; MUID:87083499; PMID:2432069
 A:Accession: A26118
 A:Molecule type: DNA
 A:Residues: 'M', 25-379; 'R', 381-472; 'S', 474-734 <REG>
 A:Cross-references: EMBL:J02638; NID:g147744; PIDN:AAA83905.1; PID:g551833
 R:Evans, S.; Dennis, P.P.
 Gene 40, 15-22, 1985
 A:Title: Promoter activity and transcript mapping in the regulatory region for genes end

A:Reference number: A23984; MUID:86137413; PMID:3005122
 A:Accession: B23984
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 'M', 25-85 <EVA>
 A:Cross-references: GB:M14425; NID:g147747; PIDN:AAA24596.1; PID:g147749
 A:Note: the authors translated the codon GCC for residue 74 as Arg
 C:Genetics:
 A:Gene: pnp
 A:Map position: 69 min
 A:Start codon: TTG
 C:Superfamily: polyribonucleotide nucleotidyltransferase alpha chain
 C:Keywords: nucleotidyltransferase
 P:25-734/Product: polyribonucleotide nucleotidyltransferase alpha chain #statue predicte
 Query Match 32.4%; Score 1152; DB 2; Length 734;
 Best Local Similarity 39.9%; Pred. No. 1e-64;
 Matches 269; Conservative 118; Mismatches 252; Indels 36; Gaps 11;
 Qy 35 GNRKLEISSGKLAFADGSAVVGSDTAVMTAVSKTSPSPQ-FMPLVVDYRQKAAAG 93
 Db 35 GQHTVLTETGMALQATAVAVVSMDDTAVFTVVGQKAKGQDPFLTVNYQERTYAG 94
 Qy 94 RLPNTYLRREVTSQDKELTSRIIDRSIRPLFPAGYFYDTQVLCNLAADVNEBDVLA 153
 Db 95 RLPGSFFRREGRPSEGETLIRLIDRPIRLPPEGFVNEGVIVATVAVSNPQVNDIVAM 154
 Qy 154 NGASVALSLSDIPFPMNGPVAGVIGIDGEEVYVPTKEMSSSTLNVAGAPKSOIVLE 213
 Db 155 IGASVALSLSDIPFPMNGPVAGVIGIDGEEVYVPTKEMSSSTLNVAGAPKSOIVLE 213
 Qy 214 ASAEMLIQQDFCHAIKVGKYYTQOIIOGIQOLVKETGKTRPPQKLFTRSP--EIVYKT 270
 Db 214 SEQQLSLSDIPFPMNGPVAGVIGIDGEEVYVPTKEMSSSTLNVAGAPKSOIVLE 269
 Qy 271 HGLAMERL---YAVFTDYHDYKVSDEAVNKRILDTBEOLKEKFPADYEIIESFNVA 327
 Db 270 AALAEARLSDAYRI---TDKERYAQVDYKSEITATLAEDELTENELGEITHAIE 324
 Qy 328 KEVRSITLYNEVKRCQGRDGLTSLRNVSCEYDMFKTLHGSALFQGGQTVLCTVPDLSLS 387
 Db 325 KNVARSVTLAGEPRIDGKEMIRGLDVRQTVLPRTHGSALFTGETQALVTATLGT--- 381
 Qy 388 GIKSDQVTAINGIKDKNFMHYEPFPVATNEIGKVTGLNREELGHGALAKALVVPVIR 446
 Db 382 ARDAQVLDLMEGETDTFLPHNFPYVSGTGVSGPKREELGHGALAKALVAVPD 440
 Qy 447 RDPFTTIVTSEVLESNSSSMASACGSLALMDSGVPISSAVAGVAGVIGVTKTDPK 505
 Db 441 MDKFPYTVRVVSEITESNGSSSMASVCGASLALMDAGVPIKAAVAGIMGLVKEGD---- 496
 Qy 506 EIBDRLRLTDLIGEDYNGDMDFKTAGTKGITALQADIKLPGPIKIVMEALQOASVAK 565
 Db 497 ---NVVVLSDILGEBDHLGDMDFKTAGTKGITALQADIKLPGPIKIVMEALQOASVAK 553
 Qy 566 KEILOINMTKISPRASRKNRPNVETVQPLSKAFVPGGYMLKKLOETGTTISQV 625
 Db 554 LILGVEQALNAPRGDISFAPRIHTIKIPDKIKVIGGGSVIRLLEIGTTIIE 613
 Qy 626 DEETFSVAPTPSVWHEARDFITEICKDQEQOLBEGAVYTTATTEIRDTGVMVLYPNM 685
 Db 614 DDGTYKIAATGEGAKAKAIRIBEI-----TABIEVGKVTYKATRYVDGAFVAGGK 668
 Qy 686 TAVLHNTQLDNERL 700
 Db 669 EG-LVHISQIADKRV 682
 RESULT 13
 AC3497
 polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) [imported] - Brucella melitensis
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002

C/Accession: AC3497
R:DelVecchio, V.G.; Kapetral, V.; Redkar, R.D.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
R.; Mazur, M.; Goldman, B.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Levese
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
A:Reference number: AD3252; PMID:11756688
A:Accession: AC3497
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-714 <KUR>
A:Cross-references: GB:AE008917, PIDN:AA53142.1, PID:G17964011, GSPDB:GN00190
A:Experimental source: strain 15M
C:Genetic: BME11961
A:Map position: 1
C:Superfamily: polynucleotide nucleotidyltransferase alpha chain
C:Keywords: nucleotidyltransferase

Query Match 32.3%; Score 1150; DB 2; Length 714;
Best Local Similarity 39.0%; Pred. No. 1.3e-64;
Matches 268; Conservative 124; Mismatches 247; Indels 48; Gaps 14;

30 VAUDELNFKLEISSGKLARFADGSAVVGSDTAVVAVTAVSKTKPSQ-FMPLVVDYRQKAAG 88
7 VEIEWGRPLTETGKIARQADGAVLATYGTAVLATVSAKBPFGODFPPLTVYQEK 66
89 AAAAGRIPTNYLRREVTSDEKILTSRIIDRSIRPLPAGYFYDQVLCNLAADVGVNEP 148
67 TYAAGKIRIGYFKKGRSENEVLVSRILDRIRPLFDGKNDQVITVQLQDLENNP 126
149 DVLAINGASVALSLSDIPNNGVAVRIGIIDGEVYVNPTRKEMSSSTLNVAGAPKQ 208
127 DILSNVASSALTTISGVFPMGPISGARVGYIDGEVLPNIDEMESLTDLVAGTSA- 185
209 IVMLBASAKNLQDFCAIKVYVYQIIQGIQOLVKEVGTGVTQKLTSPSEIYK 268
186 VLWSESEKQELPEDVMLGAVMFGHSGFQVIDAI--IKLABVAKERPRD--QPEDLS 239
269 YTHKLAMRLYAVFTDYEH----DKVSRDEAVNKRILDTBEQLEKEPE-----AD 315
240 ---ELEAKVLAVENDLREANKYTEKQARYAAVDAKAKABEHF--PEGEVEETMSAE 293
316 FYEIESFNVAKEVFRSIVNEVYKRCGRDLTSLRNVSCVDMEKTLHGSALEFORGOTQ 375
294 QPARI--FKHQAQKIVRNILDTGNRIDGRDLSVTRPIVSEVGLIPRTHGSALTFRBGTQ 351
376 VLCTVTFPSLSGSKSDOVITAINCIKDKNFMHYEPFPYATNEIGKVTGLNRRELHGA 435
352 AIVAVATLGTGE---DEQMDIALTGTYESFPLHNPFPYSVGTGRMGSGRREIGHGK 407
436 LAERKALYVIP--RDPEFTIRVTSVLSNGSSMAASCGSLALMDSGVPISSAVAGA 493
408 LAMRAIHMLPAEQFPYTIKAVSEITSSNGSSNATVCGTSLAMDGVPIVPRVAGIA 467
494 IGLVTKTDPKGELEDYRLTDLIGIEDYNGDMFKIAGTKGITALQADIKLPGIPIKI 553
468 MGLTK-----EGRRFAVLSDILGDEHDHGMDFKAVAGTEGIGTSLMDIKIGITBEI 520
554 VMEAIQOASVAKKELIQIMNTKISKPRASREKNGEVVETVQVPSKRAKFPVGGYNLKK 613
521 MKVLEQAKGRVHLIGEMAKAIISSSRALGEPARIEVMNIPTDKIDVIGSGGKVIKRE 580
614 LQAEVGTISQVDETFSVFAPTEPVMEARDFTIEICKDQEQLEFGAVYTAITEIR 673
581 IVEKIGAKINIEDDTVKIASSNGKEIAPAKWHISVAEP-----EVGEIYEGIVAKTA 635
674 DTGVVAVKLYPMNTAVLHNTQDLNERL 700
636 DFGAFVNFGRPRDG-LVHISQLAADRV 661

polynucleotide nucleotidyltransferase (EC 2.7.7.8) [imported] - *Yersinia pestis* (str
C:Species: *Yersinia pestis*
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C:Accession: AC0424
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, I
Nature 413, 523-527, 2001
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0424
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-705 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92719.1; PID:G15981414; GSPDB:GN00175
C:Genetic: BME11961
A:Map position: 1
C:Superfamily: polynucleotide nucleotidyltransferase alpha chain
C:Keywords: nucleotidyltransferase

Query Match 32.2%; Score 1146; DB 2; Length 705;
Best Local Similarity 40.0%; Pred. No. 2.2e-64;
Matches 269; Conservative 113; Mismatches 259; Indels 32; Gaps 10;

35 GNRKLEISSGKLARFADGSAVVGSDTAVVAVTAVSKTKPSQ-FMPLVVDYRQKAAG 93
12 GQHTVITETGMARQAATAVAVMSMDTAVVAVTAVGQAKAPGSGFPFLTVNYGRTYAA 71
94 RIFPNVYLRREVTSDEKILTSRIIDRSIRPLPAGYFYDQVLCNLAADVGVNEPDLAI 153
72 RIFPSFRREGRESEGGTILSRILDRIRPLPFSFNEVQVATVAVSVQNPIDVAL 131
154 NGASVALSLSDIPNNGVAVRIGIIDGEVYVNPTRKEMSSSTLNVAGAPKQIYMLE 213
132 IGASVALSLSGIPNNGTIGARVGFINDQVYVLPNTDELKESRLDVAAGT-AGAVLWVE 190
214 ASANILQODFCAIKVYVYQIIQGIQOLVKEVGTGVTQKLTSPSEIYKTH-- 271
191 SEADILSEEGMLGAVVGHGQGVVIEINAVLAEAQKPMQND--AEVMEALHAR 245
272 --KLAMERLYAVFTDYHDYKSRDEAVNKRILDTBEQLEKEFPADPEYIEESNVYAKE 329
246 VAEIABEARLDGAYVITE--KQERYQVDAIKADVTALAAQDITLDAEIQDILASVEKN 303
330 VFRSIVNEVYKRCGRDLTSLRNVSCVDMEKTLHGSALEFORGOTVLTVPDLSGI 389
304 VVRSRVLRGEPRIDREKDKMIRGLDVRTGILPRTGSALETTKGTOLVTAITLGTARDAQ 363
390 KSDQVITAINCIKDKNFMHYEPFPYATNEIGKVTGLNRRELHGALEAKALYVIP--R 447
364 NIDELM-----GERDPSFLAHYNNPPYCVGRTGAVGSPKREIGHGRALKGVLAVMSAS 419
448 DPEFTIRVTSVLSNGSSMAASCGSLALMDSGVPISSAVAGVAILGVTKDPKGEI 507
420 EFPYTIIVVBEITSSNGSSMAASVCGSLALMDAGVPIKAVAGIAMGLVKEGQ----- 473
508 EDVYLLDIIIGIEDYNGDMFKIAGTKGITALQADIKLPGIPIKIYMEAIQOASVAKKE 567
474 -NFVLSLDILGDEHDHGMDFKAVAGSDGYTALMDIKIGITREINQVLAHQAKRLH 532
568 ILQIMNTKISKPRASREKNGEVVETVQVPSKRAKFPVGGYNLKKLQAEVGTISQVDE 627
533 ILGVMGAIISTPRDISEFPARITYMKINPEKIKDVIKGGSSVIRALTDERTGTTIEED 592
628 ETFSVFAPTESVMEARDFTIEICKDQEQLEFGAVYTAITEIRDTGVAVKLYPMNTA 687
593 GTTIAATDQDKAHAIRIEI-----TAIEVGRYAKKVRIVDFGAFVAILGGKEG 647
688 VLHNTQDLNERL 700
648 -LVHISQIADRV 659

RESULT 15
G83950
polynucleotide phosphorylase pmuA [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: G83950
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: G83950

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-704 <STO>
A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA06126.1; GSPDB:GN00

A:Experimental source: strain C-125
C:Genetics:

A:Gene: pmuA
C:Superfamily: polynucleotide nucleotidyltransferase alpha chain

Query Match 32.2%; Score 1145; DB 2; Length 704;
Best Local Similarity 39.3%; Pred. No. 2, 6e-64;

Matches 265; Conservative 132; Mismatches 251; Indels 26; Gaps 11;

QY 31 AVDLGNKLEISSGKLARPADGSAVQSGDTAVMTAVSKTRPSQFMPLVVDYRQKA 90
DB 9 SIDMAGRKLTVEGTQLAQANGAVLVYRGDTAVLSTATASKEPKDLPFFPLTVNYEERLY 68
QY 91 AAGRIPTNLRREVGTSDEILTSRIIDNSIRPLPAGFYDTOTVLCNLAVDGVNEPDV 150
DB 69 AAGKIPGGFKRGESEKRAIILASRLIDRIPLPPEGFRNEVOYISVMSVDOCSSSEM 128
QY 151 LAINGASVALSTSDIPMNGPVGAVRIGIDGVEVVPTRKEMSSSTLNLVAGAPKSOIV 210
DB 129 AAMVSSIALSTSDIPFESPIAGVTYGRIDGGVINFIDQLEKSDIHLVAGT-KDALN 187
QY 211 MLEASAEMLIQDFCHAIKVGKYYTOQIIQGIQOLVKEGTVKRTPOKLFPSPIVXYT 270
DB 188 MVEAGAEVPEDEVMLEAIFGHNELKRIEPOEKIAAEVGSK-TDVVLKQYDPMLEGEV 246
QY 271 HCLAMERL-YAVFTDYEDKVRDEAVNKIRLDTEQLKEKPEADPYHIESFNVAKE 329
DB 247 RYKAEEDLQAVQVBEKHARODALEAVMDKVLTEYED-NEDVPLSEVNEILHK---IVKE 302
QY 330 VERSIVLNEYKRCDDGDLTSLNVSCEVDMFTLGSALFORGOYVLTCTVPDSLESIG 389
DB 303 EYRRLITVEKIRPDGEIDEIRPLSQVGILPRTHSGGLFTRGQTOALSICTLGAL---- 358
QY 390 KSDQVITAINGIKDKNFMLHYEFPYPATNEIGKVTGLNREIRLGHALAEKALYPVIP-R 447
DB 359 GVVQILDLGIGIESKRFMHYNFPQSVGETGPIRGPRREIGHGLGERALEPVIPISEQ 418
QY 448 DPEFTIRVTSVYLESSNGSSMASACGGSIALMDSGVPISSAVALGLVTKTDPKGEI 507
DB 419 DEPYTIRLVSEVLESNGSTSQASICASTLAMMDAGVPIKAPVAGIAMGLV-KOD----- 471
QY 508 EBYRLITDLIGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIWEALIOQASVAKKE 567
DB 472 EHVSVLTDIQEMEDALGDMDPFYAGTRKGVTAQMDIKISGIDRALLEQALQARKGRMI 531
QY 568 ILQINNKTIKSPRASKEKNGPVVEVTVQVPLSKRAKFGPGYNLKKLQALGTGTTISOVDE 627
DB 532 ILDNMLEAISBSRSELSPPAKILMTINPDKIRDIVIGSGMKINKIIEDTGKIDIED 591
QY 628 EFTSVFAPTPSVMEARDFITEICQDOEQOLEFGAVYATITTEIRDTGVMMKLYPNMTA 687
DB 592 GTIYISSADTNMNNKARIEIDIVR-----EVEVGOMYIGTVARIEKFGAFELFKGKDG 646
QY 688 VLLHNTQLDNERLN 701
DB 647 -LVHISQLAEKRVN 659

Search completed: January 8, 2004, 10:40:52
Job time : 29 secs


```

RA STRUCTURE BY NMR OF 617-692.
RX MEDLINE=97160844; PubMed=9008164;
RA Bystroff M., Hubbard T.U., Proctor M., Freund S.M., Murzin A.G.;
RT "The solution structure of the S1 RNA binding domain: a member of an
RT ancient nucleic acid-binding fold.";
CC Cell 88:235-242(1997).
CC -1- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
CC polyribonucleotides processively in the 3' to 5' direction.
CC Involved in the RNA degradation, a multi-enzyme complex important
CC in RNA processing and messenger RNA degradation.
CC -1- CATALYTIC ACTIVITY: {RNA} (N+1) + phosphate = {RNA} (N) + a
CC nucleoside diphosphate.
CC -1- SUBUNIT: Homotrimer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- INDUCTION: In response to low temperature.
CC -1- SIMILARITY: Contains 1 KH domain.
CC -1- SIMILARITY: Contains 1 SH motif domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; J02638; AAA83905.1; -.
CC EMBL; U18997; AAA57967.1; ALT_INIT.
CC EMBL; AE000397; AAC76198.1; ALT_INIT.
CC EMBL; X00761; CA25332.1; -.
CC EMBL; M14425; AAA24596.1; -.
CC PIR; H65106; H65106.
CC PDB; ISRO; 01-APR-97.
CC SWISS-2DPAGE; P05055; COLI.
CC ECODBASE; C078.0; 6TH EDITION.
CC EcGene; BG10743; pmp.
CC InterPro; IPR001247; 3_ExoRNase.
CC InterPro; IPR004087; KH_dom.
CC InterPro; IPR004088; KH_type_1.
CC InterPro; IPR000110; Ribosomal_S1.
CC InterPro; IPR003029; S1.
CC Pfam; PF00013; KH; 1.
CC Pfam; PF00726; RNase; 1.
CC Pfam; PF01138; RNase_PH; 2.
CC Pfam; PF03725; RNase_PH_C; 2.
CC Pfam; PF00575; S1; 1.
CC PRINTS; PR00681; RIBOSOMALS1.
CC SMART; SMO0332; KH; 1.
CC SMART; SMO0316; S1; 1.
CC PROSITE; PS50084; KH_TYPE_1; 1.
CC PROSITE; PS50126; S1; 1.
CC TRANSFERASE; Nucleotidyltransferase; RNA-binding; 3D-structure;
KW Complete proteome.
KW DOMAIN 557 591 KH.
FT DOMAIN 557 591 KH.
FT CONFLICT 357 590 ST MOTIF.
FT CONFLICT 450 450 G -> R (IN REF. 1).
FT CONFLICT 621 622 L -> S (IN REF. 1).
FT TURN 624 633
FT STRAND 634 635
FT TURN 636 640
FT STRAND 648 648
FT TURN 651 652
FT STRAND 651 652
FT TURN 662 665
FT HELIX 668 669
FT TURN 671 678
FT STRAND 681 682
FT TURN 682 689
FT STRAND 686
FT STRAND 689
SQ SEQUENCE 711 AA; 77101 MW; 785B7D54716FC2DE CRC64;

Query Match 32.5%; Score 1155; DB 1; Length 711;
Beet local Similarity 39.9%; Pred. No. 1e-62;
Matches 269; Conservative 118; Mismatches 252; Indels 36; Gaps 11;

```

QY	35	GNKKEISGGKALRPADSGAVVQSDVTVMVTAASKTRKPSPO--FMYLVYDYGKAAAG	93
Db	12	GQHTVLTLEGMARQATTAAMVSMMDIAVFTVVGQKAKGQDFPFLVYVIGERTTAAG	71
QY	94	RIFTNYLAREVGTSDKEILTSRIIDRSIRLPFPAQFYDTQVLCNLALVDGVEPDLAT	1533
Db	72	RIGGSPFREGRPSSEETLIARLIDRPIRPLFPEGFVNEGVIAVSVVQVNPVIDAM	1311
QY	154	NGASVALTSISDIPMNGPVAVAIIGIIDEVYVNPFRKEMSSSTLNVAAQPSQIYMLE	2133
Db	132	IGASALSTSIGLPIFNGPPIGAARVGINQVYALNPQDELKESKJLDLVAGI--EAAVLAVS	1900
QY	214	ASAKNILODFCHAIIVGVKTYQOIIIGIIOOLVETGVTYKTRPOKLFTPSP---EIVKYT	2707
Db	191	SEKQLSEDOMLGAVFEGHQOQVVIQININELVEAG---KPRMDQPEVNEALNARV	2466
QY	271	HKLAMERL---YAVFTVDYEHDKVSDAEVANKIRLDTESQLEKEFPEADPYIIIESFNVA	3427
Db	247	AALAAERLSDAYRI-----TDKQERYAQVDVAKSETIATTLAEBETLDENEIGILHAIE	3011
QY	328	KEVFRPILVNEYKRCQDRDLTSLRNVSCSDVMFKTLHGSALFQSGQTOVLCTVTFDSLES	3837
Db	302	KNVRSRVLAGBPRIDGRKKMIRLDRVRTGLPRTHSALFTGRTQALVTATLTG--	3586
QY	388	GKSDQVITAINGCIKDKFMFLHYEPFPATNIEIGVTYGLNRELRGHALAEKALPYIP-	4466
Db	359	-ADDAQVLDLMGERDITPLFHYNFPYSVGETGVSGSPKRRELGHRILAKRGVLAMPD	4177
QY	447	-RDPPTINVTSEVLSENGSSSSMASACGSLALMDSGVPISSAVAGVATGLVTKTDEKG	5050
Db	418	MDKFPYTVAVVSEITSENGSSSSMASVACGSLALMDAGVPIKAAVAGIAMGLIVEGD---	4733
QY	506	EIEDVYLLDIIIGIEDYNDMDPFKAGNNKGTLLQDITKLGPIKIVMEALQOASVAK	5655
Db	474	--NYVVLSDIIGDEHDLGDMDFKVAGSRDGISALQMDIKETGKIEMOVALQOAGAR	5301
QY	566	KEILQIMNKTIISKPRASKENGPVVEVQVPLSKRAKFEVGGYNLKKLQALTEGVITISQV	6253
Db	531	LHLTGMEQAINAPRGDISFPARPHITIKINPDKIKOVIKGGSVINHALREJOTTTIEE	5900
QY	626	DEETFSVAPPTPSVWHEARDPTTEICKDQDQOLEFGAVVYTAITTEIRDTGVMKLYPNM	6655
Db	591	DDGTVAIXIATDGEKAKHAIIRIIEI-----TAEIEVGKVYTGKATRIVDGFAFALGGK	6454
QY	686	TAVLHNTQLDNERL 700	
Db	646	EG-LVHISQIADKRV 659	
RESULT 2			
PMP_BUCAP			
ID	PMP_BUCAP	STANDARD;	PRT; 707 AA.
AC	Q8K3H5;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide phosphorylase) (PNPase).		
DE	PNP OR BUSG361.		
GN	Buchnera aphidicola (subsp. Schizaphis graminum).		
OS	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Buchnera.		
OX	NCBI_TaxID=98794;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=22084549; PubMed=12089438;		
RA	Temas I., Klasson L., Canback B., Naeslund A.K., Eriksson A.-S.,		
RA	Werngren J.U., Sandstrom J.P., Moran N.A., Andersson S.G.B.;		
RT	"50 million years of genomic stasis in endosymbiotic bacteria.";		
RL	Science 296:2376-2379(2002).		
CC	-1- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded polyribonucleotides processively in the 3' to 5' direction.		

	RESULT 2		
	PNP_BUCAP		
ID	_PNP_BUCAP	STANDARD;	PRT; 707 AA.
AC	Q8K9H5;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Polyribonucleotide nucleotidyltransferase [EC 2.7.7.8] (Polynucleotide phosphorylase) .		
DN	PNP OR BUSG361.		
CC	Buchnera aphidicola (subsp. Schizaphis graminum) .		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Buchnera.		
OX	NCB1_TaxID=98794;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
EX	MEDLINE=22084549; PubMed=12089438;		
RA	Tamas I., Klasson L., Canback B., Neeslund A.K., Eriksson A.-S.,		
RA	Werngren J.U., Sandstroem J.P., Moran N.A., Andersson S.G.E.;		
RT	"50 million years of genomic stasis in endosymbiotic bacteria." ;		
RL	Science 296:2376-2379(2002).		
CC	-I- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded polyribonucleotides processively in the 3' to 5' direction.		

CC Involved in the RNA degradingosome, a multi-enzyme complex important
CC in RNA processing and messenger RNA degradation (By similarity).
CC -1- CATALYTIC ACTIVITY: {RNA} (N+1) + phosphate -> {RNA} (N) + a
CC nucleoside diphosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Contains 1 KH domain.
CC -1- SIMILARITY: Contains 1 S1 motif domain.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL, AE014112; AAM67914.1; -
DR InterPro; IPR001247; 3_ExoRNase.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR000110; Ribosomal_S1.
DR InterPro; IPR003029; S1.
DR Pfam; PF00013; KH; 1.
DR Pfam; PF03726; PNPase; 1.
DR Pfam; PF01138; RNase_PH; 2.
DR Pfam; PF03725; RNase_PH_C; 2.
DR Pfam; PF00575; S1; 1.
DR PRINTS; PR00681; RIBOSOMAL_S1.
DR SMART; SM00322; KH; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PS50084; KH_TYPE_1; 1.
DR PROSITE; PS50126; S1; 1.
DR TRANSFERASE; Nucleotidyltransferase; RNA-binding; Complete proteome.
FT DOMAIN 553 612 KH.
FT DOMAIN 622 690 S1 MOTIF.
SQ SEQUENCE 707 AA; 78389 MW; 9D53675E6591520A CRC64;

Query Match 32.4%; Score 1152; DB 1; Length 707;
Best Local Similarity 38.5%; Pred. No. 1.5e-62;
Matches 264; Conservative 128; Mismatches 237; Indels 56; Gaps 14;

QY 35 GNRKLEISSGKLAEPADGSAVVOGSDTAVMTAV-SKTPSPSPSPMPVVDYRQKAAAG 93
DB 12 GGHITLLETGIMARQATAVMASSMDTAVFTVVGKTKNSQKFFPLTVNGERTAVG 71
QY 94 RLPNTYLRREVSTSKKIIITSRIRLPFPAGFYDTQVLCNLAVDGNEDVLA 153
DB 72 RLPGEFFREGRSENEIITARLIRPLRPKGFCEIQTIAVVSVPQINDIISI 131
QY 154 NCASVALSISDI PMNGPVGAVRIGIIDGVYVNPTRKMSSTLNVAVGAPKSIWLE 213
DB 132 ICASNALISISGIPFGPIGAARVGNVNOYVNPITIDMKSSFLLVVSQT-QNVLAVE 190
QY 214 ASAENILQDFCHAKVGVKTYQOIIQGIQOLVKEGTGVTKRPQKLFPSPEIVYKTH- 271
DB 191 AASKYLSKDKILGALMGHQGOQVIVNIRLSNNA-----SKL-----PILISYPPIN 239
QY 272 -----KLA--MERLYAVTDEYHDKVSHDEAVNKRILDTESQLEKPEP--ADPY 317
DB 240 TELERKITLAEKDISNMYLIF-----NKQRYEKLNFKAEIILKLFSENSINIS 291
QY 318 EIESFNVYAKVEPFSIVLNEKRCGRDLTSLRVSCEVDMFKTLHGSALFQRCQTVL 377
DB 292 EIEDFPEKLEKNIIVKRILNENRIRIDGRKMDIRLADIRTVGLPPTHSSLPTRGETSL 351
QY 378 CTVTEDSLESIGKSDQVITAINGIDKNKFMILYEPFPAVTEIRIGVYGLNRELGHALA 437
DB 352 VSVTIGTSRDQNDLDEL-----GDTIDNPLFHYNPPYSVSGIIGVSPKREIGHGLA 407
QY 438 EKALPYVPIR--DPEFTIRVTSVLEESNGSSSMASACGSLALMDSGVPISAVAGVAIG 495
DB 408 KSLSLAVMPKLDDEPFTYIRIVSEITENSGSSSMASVCSALMDAGVPIKSAVAGISMG 467

QY 496 LVTKTDPEKGEIEDYRLTLIDIGEDYNGDMDFKAGTNKGTALQADIKLPGIPKIVM 555
DB 468 LVKEED-----KVLSDILGDEDDHGDNDFKVSGTEBGTALQMDIKIGITNEIKR 520
QY 556 EAIQOASVAKKEIIOINKNTISKPRASRENGPVVETQVPLSKRAKVPQGVNLTQLQ 615
DB 521 IALNKAASARLIIINVMQALSKPRNEISEFAPRIHKIKINDEKIDQVIGKGSVIRMLT 580
QY 616 AETGTTISQVDEFTSVAPRPSVWHEARDPTEICKDOEQLEFGAVYRTTIEIDT 675
DB 581 BETGTIIEIDGTIKISA--TIGEKAKNAIRI--EITAEIEVGRIVSGKTVRIYDF 635
QY 676 GVMVLYENMTAVLILNTQDNERL 700
DB 636 GAFISIGIKGEG-LVHISQISNRY 659

RESULT 3
PNP_RICPR
ID PNP_RICPR STANDARD; PRT; 745 AA.
AC Q9ZDA3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Polynucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide
DE phosphorylase) (PNPase).
GN PNP OR RP504.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.B., Zomorodipour A., Andersson J.O.,
RA Sickeritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140 (1998).
CC -1- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
CC polynucleotides processively in the 3' to 5' direction.
CC Involved in the RNA degradingosome, a multi-enzyme complex important
CC in RNA processing and messenger RNA degradation (By similarity).
CC -1- CATALYTIC ACTIVITY: {RNA} (N+1) + phosphate -> {RNA} (N) + a
CC nucleoside diphosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Contains 1 KH domain.
CC -1- SIMILARITY: Contains 1 S1 motif domain.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL, AJ235272; CAA14956.1; -
DR PIR; B71654; B71654.
DR HSPD; P05055; ISRO.
DR InterPro; IPR001247; 3_ExoRNase.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR003029; S1.
DR Pfam; PF00013; KH; 1.
DR Pfam; PF03726; PNPase; 1.
DR Pfam; PF01138; RNase_PH; 2.
DR Pfam; PF03725; RNase_PH_C; 2.
DR Pfam; PF00575; S1; 1.
DR SMART; SM00322; KH; 1.

DR SMART; SM00316; S1; 1.
 DR PROSITE; PS50084; KH TYPE_1; 1.
 DR PROSITE; PS50126; S1; 1.
 KM Transferase; Nucleotidyltransferase; RNA-binding; Complete proteome.
 FT DOMAIN 558 592 KH.
 FT DOMAIN 623 691 SI MOTIF.
 SQ SEQUENCE 745 AA; 82339 MW; 885965045E76D2F2 CRC64;

Query Match 32.1%; Score 1140.5; DB 1; Length 745;
 Best Local Similarity 37.8%; Pred. No. 8.2e-62;
 Matches 255; Conservative 143; Mismatches 236; Indels 41; Gaps 12;

QY 39 LEISGKLARPADSAAVQSGDTAVMTAVSKTPSPS-QEMPLVVDYRQKAAAGRIPT 97
 DB 16 LEISGKLARQANAAVTKMGNLSILCTCVANVKKQDGGFPLINREMAYSAGKIPG 75
 QY 98 NYLREVTSTKEITLSTRIIRSLRPLFPAGYFDTQVLCMLAVDGVNEDVLAINGAS 157
 DB 76 GPFKREGKASREILVSRILDRPRLPFOAFMEHTHTCSLVDPATPVDILAIIGAS 135
 QY 158 VALSLDIPMNGPVAGAVRIGIIGDEVVNPTRKEMSSSTLNVAVAGAKSQIWMLEASAE 217
 DB 136 AALISPPAYEIVAAKVGKLGNGFVLPNLEILKTSQDLVVAAGTDS-VMMVESAH 194
 QY 218 NILQDFCHAIKVGKTYQOIIOGIDQVKEGTGKTPPKLTPSPSEIVYTHKLAMER 277
 DB 195 LLSSEDKMLEAVKFGPESFQVYIKILKELAKKPKFEMQDLY-PS-----SLKKEIK 247
 QY 278 LYAVFTYEHDKVSRDEAVNKRILDTBEQL-----KEKPEADPIEITSESNVAKKE 329
 DB 248 LPTKEVEAFKIKSKQERSTDLALYEKVLTHFVADENKRY---NNQISALAKASAD 304
 QY 330 VRSIVLNEYKRCGRDLTSLRNVCCEVDFKTLHSGALFORGOTQVCTVTF-DSLESG 388
 DB 305 ILRNKILKRNIRIDRSTTDIQLACEVGLPNSAGSLFTRGETQSLVSTTFGSLD- 362
 QY 389 IKSDQVTAINGIKDKNMFLHYEPFPYATNIEIGVTLGNREELGHALAERALYFVIR- 447
 DB 363 ---EQIVDSLSEYKERMFLNVIFFPYVSNEMPMKARSREVGKGLAMAINPILNK 419
 QY 448 -DPEPTIVTSFVLESGSSMAACGSGSLALMDSGVPISSAVAGVAILGYTKDPEKGE 506
 DB 420 VQFPYSIRVVAETTESNGSSSWATYCGSSSLAMHAGVPIKAPVAGIANGLVK-----E 472
 QY 507 IEDVLLLDIIGIEHYNDGMDPKIAGTKGKITALQADIKLPGIPIKIMBALIQASVAKK 566
 DB 473 SNKFAVLSDIIGDEYFGMDPKVAGTSSGITALMDIKISIDPKIITQIALBQRLRL 532
 QY 567 EILQIMNTIKSKPRASRKENGFPVETVQVPLSKRAKFGVPGGVNLKQLAETGVTISQVD 626
 DB 533 HILEGMNVISKPNSELSKNAPSSITVAKIDKIDIIIGPGKIKIKEICETSNAKIDISD 592
 QY 627 EETSVFAPTESVMEARDFTIEICKDOEQLEFGAVYATITIRDTGVNWKLYPMNT 686
 DB 593 DGTVAIVASDRDKIKIALDKIKAIAPER-----EIGEIFNGVMKVLDSGAFINYLGNKD 647
 QY 687 AVLNTOLDNERLN 701
 DB 648 G-FVHISEISDAKID 661

RESULT 4
 PNP_YEREN STANDARD; PRT; 706 AA.
 AC 034275;
 DT 16-OCT-2001 (rel. 40, Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 15-SEP-2003 (rel. 42, Last annotation update)
 DE Polytubonucleotide nucleotidyltransferase (EC 2.7.7.8) (polynucleotide
 phosphorylase) (PNPase).
 GN PNP.
 OS Yersinia enterocolitica.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=630;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98294047; PubMed=9632258;
 RA Goveide R.L.J., Huls in't Veld J.H.J., Kusters H.G., Mool F.R.;
 RT "The psychrotrophic bacterium Yersinia enterocolitica requires
 expression of pnp, the gene for polynucleotide phosphorylase, for
 RT growth at low temperature (5 degrees C).";
 RL Mol. Microbiol. 28:555-569(1998).
 CC -1- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
 CC polyribonucleotides progressively in the 3' to 5' direction.
 CC Involved in the RNA processing and messenger RNA degradation (By similarity).
 CC -1- CATALYTIC ACTIVITY: {RNA}(N+1) + phosphate = {RNA}(N) + a
 CC nucleoside diphosphate.
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Contains 1 KH domain.
 CC -1- SIMILARITY: Contains 1 SI motif domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Y10692; CAAT1697.1; ALT_INIT.
 CC HSSP; P05055; ISRO.
 DR InterPro; IPR001247; 3 ExoRNase.
 DR InterPro; IPR004087; KH dom.
 DR InterPro; IPR004088; KH_type_1.
 DR InterPro; IPR003029; SI.
 DR Pfam; PF00013; KH; 1.
 DR Pfam; PF03726; PNPase; 1.
 DR Pfam; PF01138; RNase PH; 2.
 DR Pfam; PF03725; RNase PH_C; 2.
 DR Pfam; PF00575; S1; 1.
 DR SMART; SM00322; KH; 1.
 DR SMART; SM00316; S1; 1.
 DR PROSITE; PS50084; KH TYPE_1; 1.
 DR PROSITE; PS50126; S1; 1.
 KW Transferase; Nucleotidyltransferase; RNA-binding.
 FT DOMAIN 557 591 KH.
 FT DOMAIN 622 690 SI MOTIF.
 SQ SEQUENCE 706 AA; 76284 MW; 9CD82CD348C3AD4A CRC64;

Query Match 31.9%; Score 1136; DB 1; Length 706;
 Best Local Similarity 39.8%; Pred. No. 1.4e-61;
 Matches 268; Conservative 115; Mismatches 259; Indels 32; Gaps 11;

QY 35 GNRLEISGKLARPADSAAVQSGDTAVMTAVSKTPSPS-QEMPLVVDYRQKAAAG 93
 DB 12 GQHTVTIETGMAAQAATAVAVSMDDTAVFVTVGQKAKRGQSFPLTVVYQERTYAG 71
 QY 94 RIFPNVLRREVTGSDKELIIRSLIRSLRPLFPAGYFDTQVLCMLAVDGVNEDVLA 153
 DB 72 RIPGSPFRREGRPSGEGLTSLRILDRPRLPFPSPFLNEGVAVIVAVSVNQINPDVAL 131
 QY 154 NGASVALSLSDIPMNGPVAGAVRIGIIGDEVVNPTRKEMSSSTLNVAVAGAKSQIWMLE 213
 DB 132 IGASNAISLSTGIPFNGPIGARVGFINDQVLTNTTBELKESRLDVLVAGT-AGAVLME 190
 QY 214 ASAENILQDFCHAIKVGKTYQOIIOGIDQVKEGTGKTPPKLTPSPSEIVYTH-- 271
 DB 191 SEADILSBDQMLGAVVGHGEOGVVENINLVAEAG---KPKMDWHAEP-VNEALHAR 245
 QY 272 --KLAMERLYAVFTYEHDKVSRDEAVNKRILDTBEQLKEKFPPEADPIEITSESNVAKKE 329
 DB 246 VAEIAAARLDGAYRIT--KQERYTQVDAIKADYTEALMDODTDLDAEIODILIGSVEKD 303

QY 330 VPSIVLNEYKRCGDGLTSLRNVCEDVMFKTLHGSALFORQOTVLTCTVPDLSBSGI 389
 DB 304 VVRSVLRGSEPRIDREKMDIRGLDVRTGVLPRTHGSALFTRGETQALVATLIGLRADQ 363
 QY 390 KSDQVITANGIKDKNPMHYEPPTATNEIGVYGLNRELGHGLAEKALYPIV--PR 447
 DB 364 NIDELM-----GERTDFLLHYNPPYVSBGTGVGSPKREKRGHGLAKRGVLAWPSPS 419
 QY 448 DEPFRTVTSVLSNNGSSSMASACGSLALMDSGPISAVAGVAGLVTKTDPKKEI 507
 DB 420 EEPYTRVVSSEITSSNNGSSSMASVCCASLALMDAGVPIKAAGVAGLVLK-----ED 472
 QY 508 EDRYLLTDILIGEDYNGDMDFKIAGTNKGTTLQADIKLPGIPIKIWEALQOASVAKKE 567
 DB 473 ENFVVLSDILIGEDHLDGDMDFKVASGRDITLQMDIKIGITREIMQALNOKASARH 532
 QY 568 ILQIMNKITSKPRASRKNQGVVETVQVPLSKRAKVGCGVNLKQLQAEVGYTISQVDE 627
 DB 533 ILGVEQALSTPGDISFPAPRIYTKINPEKIKDVIKGGSVIRALTDETGTTIDIED 592
 QY 628 ETPSVAFPTPSVHBEADFTIEICKDQOQLEFGAVYATLTIEIRDTGVMTKLYPMNTA 687
 DB 593 GTIKIATGDKAKAHLRIEEL-----TAEIEVNRIVAGKTRIVDFGAFVAGIGGKEG 647
 QY 688 VLLHNTQDLNERLN 701
 DB 648 -LVHISQIADKRD 660

RESULT 5
 PNP_HABIN STANDARD; PRT; 709 AA.

AC P4584;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DE Polyrubomucloic acid nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide
 DN phosphoglycase) (PNPase).
 GN PNP OR H10229.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KM20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Heddlom E., Cotton M.D.,
 RA Utechtback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
 RA Venter J.C., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA "whole-genome random sequencing and assembly of Haemophilus influenzae
 RT rd.";
 RL Science 269:496-512(1995).
 CC -1- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
 CC polynucleotides processively in the 3' to 5' direction.
 CC Involved in the RNA degradation, a multi-enzyme complex important
 CC in RNA processing and messenger RNA degradation (By similarity).
 CC -1- CATALYTIC ACTIVITY: (RNA) (N+1) + phosphate = (RNA) (N) + a
 CC nucleoside diphosphate.
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Contains 1 KH domain.
 CC -1- SIMILARITY: Contains 1 SI motif domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.itsb-sib.ch/announce/>
 CC or send an email to license@itsb-sib.ch).

DR EMBL: U32709; AAC21898.1; -
 DR PIR: E64056; E64056.
 DR HSSP: P05055; ISRO.
 DR TIGR: H10229; -
 DR InterPro: IPR001247; 3_ExtRNase.
 DR InterPro: IPR004087; KH dom.
 DR InterPro: IPR004088; KH type_1.
 DR InterPro: IPR003029; SI.
 DR Pfam: PF00013; KH; 1.
 DR Pfam: PF03726; PNPase; 1.
 DR Pfam: PF01138; RNase PH; 2.
 DR Pfam: PF03725; RNase PH_C; 2.
 DR SMART: SM00322; KH; 1.
 DR SMART: SM00316; SI; 1.
 DR PROSITE: PS50084; KH type_1; 1.
 DR PROSITE: PS50126; SI; 1.
 KW Transferrase; Nucleotidyltransferase; RNA-binding; Complete proteome.
 FT DOMAIN 552 622 KH.
 FT DOMAIN 621 689 SI MOTIF.
 SQ SEQUENCE 709 AA; 77006 MW; BD9E08EA6236C860 CRC64;

Query Match 31.9%; Score 1134; DB 1; Length 709;
 Best Local Similarity 36.8%; Pred. No. 1.9e-61;
 Matches 249; Conservative 137; Mismatches 251; Indels 40; Gaps 12;

QY 35 GNRKLEISGKILARPADSAVAVGSDGTAVMTAVSKTSPSQ--EMPIVVOYRKAAG 93
 DB 11 GQHTVLTGTGAIARQATIAVWASMDITVETVAVKAKOVKGQDFPLTVNQETTYAG 70
 QY 94 RIPTVYLRREVQTSDEKILTSRIIDRSIRPLFPAGYFYDTQVLCMLAVDGNBPDVLA 153
 DB 71 KIPGFFRKEGRSPSGEITLARIIDRPIRPLPBGFFVIEIQVAVSVNQISDVLAM 130
 QY 154 NGASVALSLSDIPWNGPVGAVRIGIIDGYVYVNPFRKMSSTLNLVAVGAKQIVMLE 213
 DB 131 IGASVALTLGVPFPGPIGAARVGFIDNQFVILNPTMAQKOSRDLVVAAGTDKA-VLWVE 189
 QY 214 ASAEIILQODFCMAIKVVKVTKTQIIGIQLVKRTGVT--RNPQLFPPSPBIYKT 270
 DB 190 SEADLTLEQMLAAVAVFHQOQVAVVEAIKEFAEAGPRWDVAPQ---PNTDLINKV 245
 QY 271 HKLMEIRLYAVFTYEHDKVSRDEAVNKIRIDTEBQLKEKPEADPYEIESFNVAKEV 330
 DB 246 KALAEARIGDAVRITE--KQARFEGIDAIKADVIAQIADBEKASEGKIVDIFTLBSQI 303
 QY 331 FRSIVLNEYKRCGDGLTSLRNVCEDVMFKTLHGSALFORQOTVLTCTVPDLSBSGI 390
 DB 304 VVRSVLRGSEPRIDREKMDIRGLDVRTGVLPRTHGSALFTRGETQALVATLIGLRADQ 363
 QY 391 SD-QVITANGIKDKNPMHYEPPTATNEIGVYGLNRELGHGLAEKALYPIV--PR 447
 DB 359 RDAQIITDLTGBRQDHLFHYNPPYVSBGTGVGSPKREKRGHGLAKRGVLAWPSPS 419
 QY 448 DEPFRTVTSVLSNNGSSSMASACGSLALMDSGPISAVAGVAGLVTKTDPKKEI 507
 DB 419 EEPYTRVVSSEITSSNNGSSSMASVCCASLALMDAGVPIKAAGVAGLVLK-----ED 472
 QY 508 EDRYLLTDILIGEDYNGDMDFKIAGTNKGTTLQADIKLPGIPIKIWEALQOASVAKKE 567
 DB 473 ENFVVLSDILIGEDHLDGDMDFKVASGRDITLQMDIKIGITREIMQALNOKASARH 532
 QY 568 ILQIMNKITSKPRASRKNQGVVETVQVPLSKRAKVGCGVNLKQLQAEVGYTISQVDE 627
 DB 533 ILGVEQALSTPGDISFPAPRIYTKINPEKIKDVIKGGSVIRALTDETGTTIDIED 592
 QY 628 ETPSVAFPTPSVHBEADFTIEICKDQOQLEFGAVYATLTIEIRDTGVMTKLYPMNTA 687

DB 592 GTVXIAVDSNAKNGRIEIVAEV-----EAGVIYKGVTRLDAGFAVAIVG 642
 OY 684 NMTAVLHNTOLDNERL 700
 DB 643 NKEG-LVHISQIAEERV 658

RESULT 6
 PNP_BACSU STANDARD; PRT: 704 AA.
 AC P50849;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Polynucleotide nucleoside diphosphate (EC 2.7.7.8) (Polynucleotide phosphorylase) (PNPase) (Vegetative protein 15) (VEG15).
 GN PNPase OR COMR.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96423178; PubMed=8825779;
 RA Luttinger A., Hann J., Dubnau D.;
 RT "Polynucleotide phosphorylase is necessary for competence development in Bacillus subtilis";
 RL Mol. Microbiol. 19:343-356(1996).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertoro M.G., Besieres P., Bolotin A., Borcher S., Boudier R., Boudier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruchet C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codan J.J., Conerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Guisepi G., Guy B.J., Haga K., Halesch J., Harwood C.R., Hentaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Ileya M., Jones L., Joris B., Karamata D., Kasahara Y., Klier-Bianchini M., Klein C., Kobayashi Y., Koester P., Konigstein G., Krogh S., Kumano M., Kurita K., Lapides A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noack M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Portolillo S., Prescott A.M., Priesen E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowka A., Serris S.U., Serris P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemura K., Takeuchi M., Tamakoshi A., Tanaka T., Teipstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vasarotti A., Viari A., Wambat R., Wedler E., Wedler H., Weitzinger T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis";
 RL Nature 390:249-256(1997).
 [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=97261879; PubMed=9108293;
 RA Coquard D., Hucacas M., Ott M., van Dijk J.M., van Loon A.P.,
 RA Hohmann H.P.;
 RT "Molecular cloning and characterization of the ribC gene from Bacillus subtilis: a point mutation in ribC results in riboflavin overproduction";
 RL Mol. Gen. Genet. 254:81-84(1997).
 [4]

RP SEQUENCE OF 1-20.
 RC STRAIN=168 / IS58;
 RX MEDLINE=97443988; PubMed=9298659;
 RA Antelmann H., Bernhard U., Schmid R., Mach H., Voelker U.,
 RA Hecker M.;
 RT "First steps from a two-dimensional protein index towards a response-regulation map for Bacillus subtilis";
 RL Electrophoresis 18:1451-1463(1997).
 CC -1- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded polynucleotides processively in the 3' to 5' direction. May be necessary for competence development in Bacillus subtilis. May be necessary for modification of the sRNA transcript (stabilization or translation activation).
 CC -1- CATALYTIC ACTIVITY: [RNA] (N+1) + phosphate = [RNA] (N) + a nucleoside diphosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Contains 1 KH domain.
 CC -1- SIMILARITY: Contains 1 S1 motif domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U29668; AAC3595.1; -;
 DR EMBL: Z99112; CAB13542.1; -;
 DR EMBL: Z80835; CAB02561.1; -;
 DR PIR: S70691; S70691.
 DR HSP: P05055; ISRO.
 DR Sublist: BG1491; pnpA.
 DR InterPro: IPR001247; 3 ExonNase.
 DR InterPro: IPR004087; KH dom.
 DR InterPro: IPR004088; KH_type_1.
 DR InterPro: IPR003029; S1.
 DR Pfam: PF00013; KH; 1.
 DR Pfam: PF03726; PNPase; 1.
 DR Pfam: PF01138; RNase PH; 2.
 DR Pfam: PF03725; RNase PH; 2.
 DR Pfam: PF00575; S1; 1.
 DR Pfam: PF00575; S1; 1.
 DR SMART: SM00322; KH; 1.
 DR SMART: SM00316; S1; 1.
 DR PROSITE: PS50084; KH_type_1; 1.
 DR PROSITE: PS50126; S1; 1.
 KM Competence; Transference; Nucleotidyltransferase; RNA-binding;
 FT Complete proteome.
 FT INIT MET
 FT DOMAIN 553 612 KH.
 FT DOMAIN 622 690 S1 MOTIF.
 SQ SEQUENCE 704 AA; 77332 MW; 0E305B6B9B0A7B07 CRC64;

Query Match 31.8%; Score 1132; DB 1; Length 704;
 Best local similarity 37.4%; Pred. No. 2.5e-61;
 Matches 251; Conservative 144; Mismatches 253; Indels 24; Gaps 10;

OY 32 VDLGNRLTSSGGLAFADGSAVVOGCDTAWWTATSKRPPSQMPLVVDYRQAAA 91
 DB 9 IDWAGRLTVEYGLAQANGAVWIRGDTAVASTATASKEPPLDPLTVNTEERLYA 68
 OY 92 AGRIPYVLRREVTSDEKILTSRIIDRSIRPLFPAGYFYDTQVLCNLAVDGVNBDVL 151
 DB 69 VGKIPGFIRKRGSPSKAVLASRLIDRPLPADOFRNBYVYISVMSVDQCSSEMA 128
 OY 152 AINGASVALSLSDIPWNGPVGAVRIGIIDEYVNVNPKRKMSSSTLNLVAGAPKSOIVM 211
 DB 129 AMFGSSIALSVSDIPFEGPIAGVGVGRIDQFIINPVDQLKSDINLVVAGT-KDAIIM 187
 OY 212 LEASAEIILODPFHAKVGVKTYQOIIQIGIQLVETGYTKRTPQCLFPPSPPIVXYTH 271
 DB 188 VEAGADEVPEIMLEAMTFGHEIKRLIAFOEIVAAVQ-KGSEIKLFIIDELNKKV 246

272 KLAMELVAVFTYEDKVSDEAVNKIRLDTBEOLKEKPEADPY-EIIESENVVAKVY 330
 Db 247 ALAEEDLLKAI--QVIEKHAREDAINEKVAVAKFEDSHEDITIKVQOLISKIYKNE 304
 Qy 331 PPSIVNEXKRCGGRDLTSLRNVSCEVDMFKTLHSGALPQRCOTVLCVTPDLSGK 390
 Db 305 VKRLIEEKVRPDGRVDQIRPLSSVGLPRTGSGTLTRGTQALSVCTIGAL---G 360
 Qy 391 SDQVTAINGIKDKNFMHYEPFPVATNIEGKVTGLNREELGHALAEKALYVIP--RD 448
 Db 361 DVQIIDLGLVESSEKRMHYHNPQFSGVETGPMRGGRGRIEIGHGALGERALVEVISEED 420
 Qy 449 PPTTRIVTSEVLESNGSSSMASACGSLALMDSGVPISSAVAGVALGVTKTPDPEKRIE 508
 Db 421 PPTVAVLSEVLESNGSTQASICASTLAMMDAGVIKAPVAGIANGLV-----KSG-E 473
 Qy 509 DRLRLDIIIGIEPYNDMPKTAGTKKITALQADIKLGPICKIYMEALQOASVAKKEI 568
 Db 474 HTVLTLDIQGMEALGDMQFKAAGTEKGVYALQMDIKIGLSREILAEALQAKGRMEI 533
 Qy 569 LQIMNKTIKSPKRSKENGPPVETVQVPLSKRAKFPVGGVYMKKQLAETGVYISQVDE 628
 Db 534 LMSMTATLESRELSRYAPKILMTINPDKIRDVIGPSGQINKIIEETGVKIDIEQOG 593
 Qy 629 TFSVPAFSPVMEADFTTEICKDDQEQLEFGAVYATITIEIRDTGVYKLYPMNTAV 688
 Db 594 TIFISSTDSGNGQAKKII-----EDLVREVEVGQYLYLKVRIEKGAFVEIFSGKOG- 647
 Qy 689 LLAHTOLDNERL 700
 Db 648 LVHISSEALERV 659

RESULT 7
 PNP_BUCAL STANDARD; PRT; 707 AA.

AC P57454;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Polynucleotide nucleic acid transferase (EC 2.7.7.8) (Polynucleotide
 phosphotransferase) (PNPase).
 GN PNP OR BU373.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 symbiotic bacterium).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 OX NCBI_TaxID=118099;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tokyo 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 Buchnera sp. APS.";
 RL Nature 407:81-86(2000).
 CC - FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
 polynucleotides processively in the 3' to 5' direction.
 CC Involved in the RNA degradation, a multi-enzyme complex important
 CC in RNA processing and messenger RNA degradation (By similarity).
 CC - CATALYTIC ACTIVITY: {RNA} (N+1) + phosphate = {RNA} (N) + a
 CC nucleoside diphosphate.
 CC - SUBUNIT: Homotrimer (By similarity).
 CC - SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC - SIMILARITY: Contains 1 KH domain.
 CC - SIMILARITY: Contains 1 SI motif domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.
 CC EMBL, AP001119, BAB13077.1; -
 DR HSSP, P05055; ISRO.
 DR InterPro, IPR001247, 3 ExonNase.
 DR InterPro, IPR004087, KH_dom.
 DR InterPro, IPR004088, KH_type_1.
 DR InterPro, IPR003029, SI.
 DR Pfam, PF00013, KH; 1.
 DR Pfam, PF03726; PNPase; 1.
 DR Pfam, PF01138, RNase; PH; 2.
 DR Pfam, PF03725, RNase; PH; 2.
 DR Pfam, PF00575, SI; 1.
 DR SMART, SMO0322, KH; 1.
 DR SMART, SMO0316, SI; 1.
 DR PROSITE, PS50084, KH_type_1; 1.
 DR PROSITE, PS50126, SI; 1.
 DR Transferrase; Nucleotidyltransferase; RNA-binding; Complete proteome.
 FT DOMAIN 557 591
 FT DOMAIN 622 690
 FT DOMAIN 707 707
 SQ SEQUENCE 707 AA; 78234 MW; 75FPAIBBE08105B7A CRC64;

Query Match 31.3%; Score 1115; DB 1; Length 707;
 Best Local Similarity 37.2%; Pred. No. 2.7e-60;
 Matches 254; Conservative 142; Mismatches 237; Indels 50; Gaps 13;
 Qy 35 GNRKLEISGKLARADSAVQSGDTAVMTAVSKTPSPSQ-FMPLVVDYRQKAAAG 93
 Db 12 GQHTTLETGYARAKAAVMSNDEIVFVTVGQKKIHGQKFPPLTVNQETTYAAG 71
 Qy 94 RPTNYLREVQTSDEKILTSRIIDRSIRPLEPAGFYDTQVLCMLAVDGVNBDVLAI 153
 Db 72 RIPGFPFRREGSPSNEILTALRIDRPLRPLEPKKFLMEIQIATVAVSNPQINDIISI 131
 Qy 154 NCASVALSLSDIPANGPVGAVRIGIIDSEYVNPFRKSSSTLNLVYAGAKSQIWMLE 213
 Db 132 IGASNALSLSGIPFYGPAVGAARVGYINQYILNPISDMKSSLDLVVSGT-QNALWME 190
 Qy 214 ASAEKILQDFCHAKVVKYTOQIIQGIQOLVKETGVTKRPOKLTFTSPSEIVY--TH 271
 Db 191 AESKILSEKILGAILIFHQOQOVVYINIRISNEA-----SKL---PVIISYPETN 239
 Qy 272 KLAMEL-----YAVFTYEDKVSDEAVNKIRLDTBEOLKEKPEADPYEII 320
 Db 240 KTLKELKINSPEKINSIDAVVIF-----NKQRIEKLINSIKENIILFLPDENSNIDTLEIE 294
 Qy 321 ESENVVAVKPSIVLNEKRCGGRDLTSLRNVSCEVDMFKTLHSGALPQRCOTVLCV 380
 Db 295 DIFQKIEKKVVKRILSNQTRIDGREKOMIALDVRTGILPRTGSGALFTGSETQSLVSV 354
 Qy 381 TFDLSGSKSDQVITANGIKDKRPMHYEPFPVATNIEGKVTGLNREELGHALAEK 440
 Db 355 TLGTSRDAQNIDELL---GGRINDFLPHYNPFPVSEIIGVSGPKRREIGHGLARLS 410
 Qy 441 LYPVIP--RDPPTTRIVTSEVLESNGSSSMASACGSLALMDSGVPISSAVAGVALGVT 498
 Db 411 LLAWVPTLENPTTRIVTSEVLESNGSSSMASVCGASIALMDAGVPISSAVAGISMGVLYK 470
 Qy 499 KTDPEKGEIEDYRLTDILGIEDYNGDMDFKTAGTKKITALQADIKLGPICKIYMEAL 558
 Db 471 -----EGNHVLLSLDILGDEHDHGDVFKVAGTEBGTITLQOMDKLEGITNBIHSL 523
 Qy 559 QOASVAKKEILQIMNKTIKSPKRSKENGPPVETVQVPLSKRAKFPVGGVYMKKQLAET 618
 Db 524 NEARLARLHILANVQALNERSSEISEPARIHIKINPEKIKDVYIGGGSVIRLTLTEET 583
 Qy 619 GTTISQVDEFTFSVPAFPPSVMEADFTTEICKDDQEQLEFGAVYATITIEIRDTGV 678
 Db 584 GTTIEIEDDGVTKI---STYKEKAKNAIRRI--KEITAEIIEVGRVIGSKYTRIDPFAF 638
 Qy 679 VKLYPNMTAVILLANTOLDNERLN 701
 Db 639 VSIQIGKEG-LVHISQISDKRVD 660

```

RESULT 8
PNP_PROLU STANDARD; PRT; 709 AA.
ID_PNP_PROLU
AC P41121;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide
phosphorylase) (PNPase) (CNP87K).
GN PNP OR PPH.
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-13.
RC STRAIN=K122;
RX MEDLINE=94266731; PubMed=8206856;
RA Clarke D.J., Dowds B.C.A.;
RT "The gene coding for polynucleotide phosphorylase in Photorhabdus sp.
RT strain K122 is induced at low temperatures.";
RL J. Bacteriol. 176:3775-3784(1994).
CC -1- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
CC polyribonucleotides processively in the 3' to 5' direction.
CC -1- CATALYTIC ACTIVITY: {RNA}(N+1) + phosphate = {RNA}(N) + a
CC nucleoside diphosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- INDUCTION: In response to low temperature.
CC -1- SIMILARITY: Contains 1 KH domain.
CC -1- SIMILARITY: Contains 1 SI motif domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
-----
CC
CC EMBL: X76069; CA53671.1; -.
CC PIR: S38883; S38883.
CC HSSP: P05055; ISRO.
CC InterPro: IPR001247; 3 ExonNase.
CC InterPro: IPR004087; KH_dom.
CC InterPro: IPR004088; KH_type_1.
CC InterPro: IPR000110; Ribosomal_SI.
CC InterPro: IPR003029; SI.
CC Pfam: PF00013; KH; 1.
CC Pfam: PF00726; PNPase; 1.
CC Pfam: PF01138; RNase_PH_2.
CC Pfam: PF03725; RNase_PH_C; 2.
CC Pfam: PF0575; SI; 1.
CC PRINTS: PRO0681; RIBOSOMALSI.
CC SMART: SM00322; KH; 1.
CC SMART: SM00316; SI; 1.
CC PROSITE: PS50084; KH_TYPE_1; 1.
CC PROSITE: PS50126; SI; 1.
CC TRANSFERASE: Nucleotidyltransferase; RNA-binding.
CC DOMAIN 553 612 KH.
CC FT DOMAIN 622 690 SI MOTIF.
CC SEQUENCE 709 AA; 76745 MW; EBRCC59FC921E68 CRC64;

```

```

Query Match 31.1%; Score 1106; DB 1; Length 709;
Best Local Similarity 38.3%; Pred. No. 9.5e-60;
Matches 258; Conservative 126; Mismatches 256; Indels 34; Gaps 12;

```

```

QY 35 GNRKLEISSGLAPADSGDTAVMTAVSKTPSPSQ-FMPVIVYRQKAAAG 93
DB 12 GQHTVTLETGMAAGQATAAVNMDDTAIVFTVVGQKVKAGQDPFLTVYQERTYAG 71
QY 94 RIPTNYLRREVGTSDKEILTIRIIDSIRPLFPAGYFYDTQVLCNLAVDGNPDLAI 153

```

```

Db 72 RIPGFFRRRGRGEGEETLVARLIDRLRLPPEGFLNEVRIATVAVSVNPQINDIVAM 131
QY 154 NGASVALSLSDIPWNGVAVRIGIIDGEVYVNPTRKEMSSSTLNIVAGAPSQIMLE 213
Db 132 IGASALALSGIPNGEIGARVGYINDQVLPNTSDLNKSRDLVLSGT-AGAVLMVE 190
QY 214 ASANILQDPFCFAIKGVYRTOOIIQGIQOLVKEVTVTRTPQKLPFPPEIVKYTH-- 271
Db 191 SEADLLEBQMLGAVVGHQOQVVIDNIALAAENKKEKMD---WPEP-VNQALHDR 245
QY 272 -KIAMERLYAVFTYEHNDVSSDEAVNKRILDTQELKEKFPADYEIIESFNVAKE 329
Db 246 VAEIAESRLDAVAYITE--KQERYAQVDAIKDEVTALALQDETLERAEIHEILGLEKN 303
QY 330 VFSIVINERYKRDGDRGLTSLRVNSGVDMPKTLHGSALEFORQOTVLCVTDSLSGI 369
Db 304 VNSRVLGSEPRIDGRKDMVRALDVRTGVLPRTGSALEPRTBETQALVTATL---GT 358
QY 390 KSD-QVTTAINGIKDKFMFLHYEPYPATNEIGKVTGLNREIGHALAEKALVPIPR- 447
Db 359 ERDAQIIDLMLGERTDRFLHYNFPYISVGETMGSPKRRREIGHGLARGLAVWPKA 418
QY 448 -DEPFTIRVTSVLENSGSSSMASACGSLALMDSGVPISSAVAGVAILVTKDPEKE 506
Db 419 NEFPYTVRVSVETESNGSSSMASVCGASLAMDAGVPIKAAVAGIMGLVKEGD----- 473
QY 507 IEDRLITDLIGEDVNDGDFKACGNKGTLOQDIKPGPIKIVMAIQASVAKK 566
Db 474 -NEVVLSDILGDBDHGDMPFVAGSCGISALQMDIKEGITREIMOVALQAKARL 531
QY 567 EIIQIMKRTSKPRASHKENGPNVEVQVPLSRKAFVGGVNLKQLQAEVTTISQVD 626
Db 532 HILSVMEQALTTTRDDISQAPRIHTIKINPDIKKIVIGKGSVIRALTBETGTTIED 591
QY 627 EETFSVAPLPSPVWHEARDPITEICKDQEQLEFGAVYATTTREIDTGVWVLYPMT 686
Db 592 DGFVKIATGEGAKHAISIRIEI-----TAETIEVGRIYAGKTVRIYDFGFAVIGGKE 646
QY 687 AVLHNTQDLNERL 700
Db 647 G-LVHISQIDAKRY 659

```

```

RESULT 9
PNP_PSEBU STANDARD; PRT; 701 AA.
ID_PNP_PSEBU
AC O87792;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide
phosphorylase) (PNPase).
GN PNP.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TMB;
RA Favaro R., Deho' G.;
RT "Identification and cloning of genes involved in RNA turnover in
RT Pseudomonas putida.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
CC polyribonucleotides processively in the 3' to 5' direction.
CC Involved in the RNA degradation, a multi-enzyme complex important
CC in RNA processing and messenger RNA degradation (By similarity).
CC -1- CATALYTIC ACTIVITY: {RNA}(N+1) + phosphate = {RNA}(N) + a
CC nucleoside diphosphate.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

```

CC -1- SIMILARITY: Contains 1 KH domain.
 CC -1- SIMILARITY: Contains 1 SI motif domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Y18132; CAA77048.1; --
 CC HSSP: P05055; ISKO.
 CC InterPro: IPR001247; 3_ExonNase.
 CC InterPro: IPR004087; KH dom.
 CC InterPro: IPR004088; KH type 1.
 CC InterPro: IPR000110; Ribosomal_S1.
 CC InterPro: IPR003029; S1.
 CC Pfam: PF00013; KH; 1.
 CC Pfam: PF03726; NPase; 1.
 CC Pfam: PF01138; RNase PH; 2.
 CC Pfam: PF03725; RNase PH_C; 2.
 CC Pfam: PF00575; S1_1.
 CC PRINTS: PR00681; RIBOSOMAL_S1.
 CC SMART: SM00322; KH; 1.
 CC SMART: SM00316; S1; 1.
 CC PROSITE: PS50084; KH TYPE_1; 1.
 CC PROSITE: PS50126; S1_1.
 CC Transferrase; Nucleotidyltransferase; RNA-binding.
 CC DOMAIN 558 592
 CC FT 623 691 S1 MOTIF.
 CC SEQUENCE 701 AA; 74984 MW; 5DB8F1A446DF0F2C CRC64;
 SQ
 Query Match 30.0%; Score 1068; DB 1; Length 701;
 Best Local Similarity 36.3%; Pred. No. 1.9e-57;
 Matches 247; Conservative 130; Mismatches 259; Indels 44; Gaps 14;
 35 GNRKLEISSGKLFAPDAGSAAVVO-SGDTAVMTAVSCTSPSPQ-FMPIVVDKRAAA 92
 11 GGSTVLTGRTARQATGAVLVVDVYLVTVGAKADPKGFPSTHYQETVAA 70
 93 GRIPNYLREVGTSDEKILTSRIIDRSIRPLEPAGFYDYOTLNLAVDGVNEEDVLA 152
 71 GKIPGFPFREGSPSEKELTSLRIDRPIPLPEEGFMNEGVCTVSTSKTIDPDIA 130
 153 INCASVALSLSDIPNMGPGAVRIGIIDE-YVNTTRKMSSTLANLYAGAPKQIYM 211
 131 MGTSAALAIISGIPFSGPIGAARVAHESGTILNFTYQOLAASSIDMVAAGT-SDAVLM 189
 212 LBSAENILQODPCHAIKYGVKTYQOIIQOIQOLVETGVTKRTPQKLPSPPEIKYTH 271
 190 VESNOELTEDQMLGAVLPAHDFQAVIQAVKELAEAG---KPTWMDKPAVANTELFN 245
 272 KLAMELVAVFTDY---EHDVKS---RDEAVNKIRLDTEBQLKEKPEADPYEILS 322
 246 AIRAEGBEAVSQCYTITVKAADRYARIGELRDQVAKF-----SGBEQGPSAS--EVEKI 297
 323 PNVVAKGVRSIYLAETKACDGRDLTSLKRVSCEDUMFTLGSALFQGCQOVLCTVTF 382
 298 FGEIERYTVRENIVNKPRIIDGNDNTVPLNIEVGLVLEKRTGSALEFTGERTQALVAVLT 357
 383 DLSLEGKSDOYVTAINGIKDNFMHYPEPPYATNIEIKVYGNRREHGHGALAEKALY 442
 358 GT---ARDAQLDLTLEGKQKPFMLAHNPPFVSVECGMGAGREIIGHGELARRSQ 413
 443 PVIIRD--PPTIRVTSVLAESNGSSMSASACGSLALMDSGVPISSAVAGVAGLVYLT 500
 414 AMLPADVFPYTRIVVSEITNESGSSMSVCGSALMDAGVPMKAPVAGIAMGLVKEG 473
 501 DEKGIERYRLTLTLGIEDNMGDMDFKLAGNKITLADIKLPGIPKIKWEALIQ 560
 474 D-----KFAVLTDLIGDEHLDGMDPFAVAGTAKAVTALQMDIKINGITEEMETALQ 526

QY 561 ASVAKKEILQIMNKTISKPRSRKENGVEVTVQVLSKRAKFGVGGYNLKKQLAETGV 620
 DB 527 ALERANTLIGGNGVIGGSRTELSSANAPTMIAKIDTKIRVDIGSGATIRACEFTVA 586
 QY 621 TISQVDETFSPVATPSPVHARPFITBICDDDEQQLFPAVATATITERDGVWVK 680
 DB 587 SIDIDBDSIKIFGTGKADAAKORILGI-----TAAEIGKIYGVVERIVDGAFVN 641
 QY 681 LYPNMTAVLHNTOIDNRL 700
 DB 642 ILPGKDG-LVHISMLSDARV 660
 RESULT 10
 ECX1 SULSO
 ID ECX1 SULSO STANDARD; PRT; 248 AA.
 AC Q9UXC2;
 DT 15-SEP-2003 (Rel. 42; Created)
 DT 15-SEP-2003 (Rel. 42; Last sequence update)
 DT 15-SEP-2003 (Rel. 42; Last annotation update)
 DE Probable exosome complex exonuclease 1 (EC 3.1.13.-).
 GN SSO0735.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_Taxid=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=20165948; PubMed=10701121;
 RA Charlebois R.L., Singh R.K., Chan-Weiler C.C.-Y., Allard G., Chow C.,
 RA Confalonieri F., Curtis B., Duguet M., Eraso G., Pagny D.,
 RA Gaasterland T., Garrett R.A., Gordon P., Jeffries A.C., Kozera C.,
 RA Kuschwa N., Lafleur E., Medina N., Peng X., Penny S.L., She Q.,
 RA St Jean A., van der Oost J., Young F., Zivanovic Y., Doolittle W.F.,
 RA Ragan M.A., Sensen C.W.;
 RT "Gene content and organization of a 281-kbp contig from the genome of
 RT the extremely thermophilic archaeon, Sulfolobus solfataricus P2.";
 RN Genome 43:116-136(2000).
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=2132296; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Awayez M.J., Chan-Weiler C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Noors A., Eraso G., Fletcher C., Gordon P.M.K.,
 RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2";
 RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 CC -1- FUNCTION: Probably involved in the 3'-5' degradation of a variety
 CC of RNA species (potential).
 CC -1- SUBUNIT: Component of the archaeal exosome multienzyme
 CC ribonuclease complex (potential).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
 CC -1- SIMILARITY: BELONGS TO THE RNASE PH FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Y18930; CAB57569.1; --
 CC EMBL: AE006698; AAK1031.1; --
 CC PIR: H90221; H90221.
 CC HAMAP: MF_00591; -; 1.
 CC InterPro: IPR001247; 3_ExonNase.
 CC Pfam: PF01138; RNase PH; 1.

DR Pfam; PF03725; RNase PH C; 1.
 KW Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.
 SQ SEQUENCE 248 AA; 27578 MM; EMB2C2C89DD5854C CIRC64;

Query Match 5.9%; Score 209; DB 1; Length 248;
 Best Local Similarity 27.2%; Pred. No. 6.4e-06;
 Matches 72; Conservative 45; Mismatches 86; Indels 62; Gaps 10;

334 VLNEYKCDGRLTSLNNSCEVDMFKTLHGSALFQGGTQVLTCTVFDLSGSKSDQ 393
 12 LILDDGKRTDGRKPELMSIKIEIGVLRKADSSALFEKNGT-----K 53
 394 VITAINIKDKN-----FMLHYEPFYATNEIGKTYGLNRRELHGALAEKALP 441
 54 AIAAYGPKEMPHRLSLPDAVLAVRYHMTPEFTDE-RKQPAERREIELSKVREALE 112
 442 YPVLPDPPT-IRYTSVLESNGSSMASACGSLALMDSGVPISSAVAGVAGLVTKT 500
 113 SAVLELPPTAIDVFTEILQADGSRVLSMAASLALADAGIPWRDLIAGVAGKA--- 169
 501 DPEKEIDYRLTDILGIEDYNGMDPKIA--GTNGITALQA-----DI 544
 170 -----DGVITLDLNETDMWGEADMPILAMPISLNO-VTLFQINSGMTPDEFROAFDL 220
 545 KLPGIP-KIYMEALQOASVAKKE 567
 221 AVKGINITVNLREALKSKYVEFKE 245

RESULT 11
 ECX1_AERPE STANDARD; PRT; 246 AA.
 AC O9YC03;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Probable exosome complex exonuclease 1 (EC 3.1.13.-).
 GN APT1447.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;
 OC Desulfurococcaceae; Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KI;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 CC -1- FUNCTION: Probably involved in the 3'-5' degradation of a variety
 CC of RNA species (Potential).
 CC -1- SUBUNIT: Component of the archaeal exosome multienzyme
 CC ribonuclease complex (Potential).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO THE RNASE PH FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AP000061; BAA80445.1; -
 DR PIR; G72623; G72623.
 DR HAMAP; MF_00591; -; 1.

DR InterPro; IPR001247; 3 Exonuclease.
 DR Pfam; PF01138; RNase PH; 1.
 KW Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.
 SQ SEQUENCE 246 AA; 26670 MM; 9548CEA2B2F4B86 CIRC64;

Query Match 5.9%; Score 208.5; DB 1; Length 246;
 Best Local Similarity 27.1%; Pred. No. 6.8e-06;
 Matches 69; Conservative 47; Mismatches 90; Indels 49; Gaps 8;

335 VLNEYKCDGRLTSLNNSCEVDMFKTLHGSALFQGGTQVLTCTVFDLSGSKSDQ 394
 11 LIRGRHDDGLPELIDPVHMQVGIHNDGSAVLEGRTRVL----- 53
 395 ITAINIKDKN-----FMLHYEPFYATNEIGKTYGLNRRELHGALAEKALP 443
 54 -AAAYGPREPHQRYVLPDAALRVRYHMAFPSTDE-RKBPASRREIELSKVREALEP 111
 444 VI-PRDPPPT-IRYTSVLESNGSSMASACGSLALMDSGVPISSAVAGVAGLVTKTD 501
 112 VLALEFPRTVIDVFLEVLQADGSTRRAVTAASLALADAGIPRALVGVAGVAGKIGV- 170
 502 PEKEIDYRLTDILGIEDYNGMDPKIAGTKN--GITALQADIKLPG----IPKIYV 555
 171 -----LVVDVDELEDMYGEADMPVAAAPDIGEITLLQLNGVLTGSEFFRTALAMAL 220
 556 EALQOASVAKKEIQ 570
 221 RAIDRVVMEKEAIR 235

RESULT 12
 ECX1_METMA STANDARD; PRT; 493 AA.
 AC Q8PRT8;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Probable exosome complex exonuclease 1 (EC 3.1.13.-).
 GN MM2623.
 OS Methanosarcina mazei (Methanosarcina frisia).
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2209;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Goel / Goel / ATCC BAA-199 / DSM 3647 / OCM 88;
 RX MEDLINE=22120827; PubMed=1215824;
 RA Deppe-meier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
 RA Martensen-Arias R., Henne A., Wierzer A., Baeumer S., Jacobi C.,
 RA Brueggemann H., Lienhard T., Christmann A., Boemcke M., Steckel S.,
 RA Bhattacharya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
 RA Fritz H.-U., Gottschalk G.;
 RT "The genome of Methanosarcina mazei: evidence for lateral gene
 RT transfer between Bacteria and Archaea.";
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
 CC -1- FUNCTION: Probably involved in the 3'-5' degradation of a variety
 CC of RNA species (Potential).
 CC -1- SUBUNIT: Component of the archaeal exosome multienzyme
 CC ribonuclease complex (Potential).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO THE RNASE PH FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AE013507; AAM32319.1; -
 DR HAMAP; MF_00591; -; 1.

CC or send an email to license@ib-sib.ch).
CC -----
DR EMBL; AB009863; AAL64027.1; -
DR HAMAP; MF_00591; -; 1. Exonase.
DR InterPro; IPR001247; 3. Exonase.
DR Pfam; PF01138; RNase PH; 1.
DR Pfam; PF03725; RNase PH C; 1.
DR Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ SEQUENCE 246 AA; 27148 MW; 301145A699D3D062B CRC64;

Query Match 5.2%; Score 186; DB 1; Length 246;
Best Local Similarity 27.6%; Pred. No. 0.0016;
Matches 61; Conservative 32; Mismatches 78; Indels 50; Gaps 7;

QY 341 RCDSGRDLSLRNVCSEVDMFTLHGSALFQSGQTOVLCTVTEDSLSGIKSDQVITAIN 400
DB 14 RADGTPQMEVKIAVGVSNADSNMVSIGAT-----TAVAAVYG 55
QY 401 -----IKDKNFM-LHYEPFPYATNIGKVTGLNRREIGHALAEKALYPV----- 445
DB 56 PREMHPRHLS/PDRGVKRVYHMA/PSTKDERKSPTRRREIEISKILREALPALVLEQ 115
QY 446 -PRPPFTIRYTSVLENSGSSSMASACGSLALMDSGVPISSAVAGVAILVTKTDPK 504
DB 116 YPRS---RIDVFVEILQADGSTRVASLTAASLALADAGIYMRDLVGVSGVIV----- 165
QY 505 GEIEDYRLTDILGIEDYNGDMDFKIA--GNNKGITLAD 543
DB 166 ----DGTIVLDLNGLEDQYGEGLDPLGYMPNLKRTLLQLD 202

RESULT 15
ECX1_PVRAB STANDARD; PRT; 249 AA.
ID ECX1_PVRAB
AC Q9V119;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable exosome complex exonuclease 1 (EC 3.1.13.-).
GN PYRAB06100 OR PAB0420.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Ozeay;
RX PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Hellig R., Lecompte O.,
Poch O., Priour D., Querellou J., Ripp R., Thierry J.-C.,
Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
CC -1- FUNCTION: Probably involved in the 3'-5' degradation of a variety
of RNA species (Potential).
CC -1- SUBUNIT: Component of the archaeal exosome multienzyme
ribonuclease complex (Potential).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE RNASE PH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ibt-sib.ch/announce/>
CC or send an email to license@ib-sib.ch).
CC -----
DR EMBL; AJ248284; CAB49532.1; -
DR PIR; E75181; E75181.
DR HAMAP; MF_00591; -; 1.
DR InterPro; IPR001247; 3. Exonase.

DR InterPro; IPR002160; Kunitz_legume.
DR Pfam; PF01138; RNase PH; 1.
DR Pfam; PF03725; RNase PH C; 1.
DR Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ SEQUENCE 249 AA; 27678 MW; ACB0748E1BF07446 CRC64;

Query Match 4.9%; Score 175.5; DB 1; Length 249;
Best Local Similarity 27.5%; Pred. No. 0.0007;
Matches 72; Conservative 40; Mismatches 93; Indels 57; Gaps 11;

QY 340 RCDSGRDLSLRNVCSEVDMFTLHGSALFQSGQTOVLCTVTEDSLSGIKSDQVITAIN 399
DB 16 RRIDGRKKYELRPIMKEVGLKNANGSAVIEWGK-----NKIIAAVY 57
QY 400 GIDKN-----FMLHYEPFPYATNIGKVTGLNRREIGHALAEKALYP-VIPR 447
DB 58 GPRELHFKHLQRPDRALIRVRIMMAPSVBE-RKKQEPDRRSIEISKVITGALLEPALILE 116
QY 448 DPEPT-IRYTSVLENSGSSSMASACGSLALMDSGVPISSAVAGVAILVTKTDPKGE 506
DB 117 MFPRTAIDVPIEVLQADAGTRVAGITPAASLALADAGIPMRDLVAACAAGKI-----EGE 170
QY 507 IEDYRLTDILGIEDYNGDMDFKIA--GNNKGITLADIKLPGLPIK-IVMEALIQASV 563
DB 171 I-----VLDINKSEEDNYGEADVPAIMPKNIDITLLQMD-----GYLTKDEFIEAVKLAIK 221
QY 564 AKKEIIQIMKNTISKPRASKE 585
DB 222 GAKAVYQ-----KOREALKE 236

Search completed: January 8, 2004, 10:39:16
Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2004, 10:31:51 ; Search time 42 Seconds
(without alignments)
4331.597 Million cell updates/sec

Title: US-09-907-907A-42

Perfect score: 3557

Sequence: 1 DGFLLPRDRALTOQLVRA.....TAVLHNTQDNERLITLP 705

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SEPREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_ricent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriopl:*
- 17: sp_archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3516	98.8	783	4	Q81WX1
2	3512	98.7	783	4	Q81WXS
3	3480	92.2	783	11	Q8K1R3
4	2484	69.8	589	11	Q8R2U3
5	2389	67.2	540	11	Q9DC52
6	2385	64.2	504	4	Q96T05
7	1958	55.0	771	5	Q9V9X7
8	1940.5	54.6	720	5	Q95RX7
9	1761	49.5	717	5	Q81H29
10	1365.5	38.4	991	10	Q987G6
11	1276	35.9	745	5	Q965N3
12	1253.5	35.2	707	16	Q8RA43
13	1204.5	33.9	703	16	Q97145
14	1198	33.7	733	16	Q8KBY3
15	1190.5	33.5	698	16	Q99UJ8
16	1188.5	33.4	698	16	Q8NMY9

17	1187.5	33.4	713	2	Q9ZAE1	Q9ZAE1 thermus the
18	1176	33.1	717	16	Q9ZSMO	Q9ZSMO rhizobium m
19	1170	32.9	703	16	Q8RIAI	Q8RIAI fusobacteri
20	1169	32.9	715	16	Q8DGM9	Q8DGM9 synchococc
21	1166.5	32.8	713	16	Q8UJ56	Q8UJ56 agrobacteri
22	1165.5	32.8	701	16	Q8CST1	Q8CST1 staphylococ
23	1163	32.7	717	16	Q8XHP6	Q8XHP6 ralsstonia s
24	1161	32.6	711	16	Q8Z310	Q8Z310 salmonella
25	1160.5	32.6	702	16	Q8XJ54	Q8XJ54 clostridium
26	1160	32.6	708	16	Q8DBU9	Q8DBU9 vibrio vuln
27	1160	32.6	711	16	Q8ZLT3	Q8ZLT3 salmonella
28	1160	32.6	721	2	Q93MR3	Q93MR3 salmonella
29	1158.5	32.6	749	16	Q92HV7	Q92HV7 rickettsia
30	1158	32.6	734	16	Q8X9M3	Q8X9M3 escherichia
31	1156	32.5	810	16	Q9RSR1	Q9RSR1 deinococcus
32	1153	32.4	718	16	P72659	P72659 synchocyst
33	1153	32.4	740	16	Q8FD87	Q8FD87 escherichia
34	1151	32.4	714	16	Q8FXS9	Q8FXS9 bruceella su
35	1150	32.3	714	16	Q8YEB7	Q8YEB7 bruceella me
36	1146	32.2	705	16	Q8ZBC6	Q8ZBC6 yersinia pe
37	1146	32.2	737	16	Q8D1D1	Q8D1D1 yersinia pe
38	1145	32.2	704	16	Q8KAE3	Q8KAE3 bacillus ha
39	1140.5	32.1	775	16	Q66593	Q66593 aquilex aeo
40	1138.5	32.0	718	16	Q8YPI1	Q8YPI1 anabaena sp
41	1137	32.0	719	16	Q9PGQ9	Q9PGQ9 xyella fas
42	1132	31.8	702	16	Q8EHL1	Q8EHL1 shewanella
43	1130	31.8	723	16	Q92C23	Q92C23 listeria in
44	1127	31.7	723	16	Q8Y7F1	Q8Y7F1 listeria mo
45	1126	31.7	715	16	Q98B13	Q98B13 rhizobium 1

ALIGNMENTS

RESULT 1	Q81WX1	PRELIMINARY:	PRT:	783 AA.
AC	Q81WX1			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	3'-5' RNA exonuclease.			
GN	OLD35.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	PubMed:12473748;			
RA	Leszczynska M., Kang D.-C., Sarkar D., Su Z.-Z., Holmes M.,			
RA	Valerie K., Fisher P.B.,			
RT	"Identification and cloning of human polynucleotide phosphorylase,			
RT	hnpase (old-35), in the context of terminal differentiation and			
RT	cellular senescence."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16636-16641(2002).			
DR	EMBL; A1027528; AAK13047.1; ..			
KW	Exonuclease.			
SQ	SEQUENCE 783 AA; 85964 MW; 912939AF55309E2C CRC64;			
Query Match	98.8%; Score 3516; DB 4; Length 783;			
Best Local Similarity	99.4%; Pred. No. 2.9e-219;			
Matches	966; Conservative 2; Mismatches 2; Indels 0; Gaps 0;			
QY	1 DGFLLPRDRALTOQLVRAVAVDIGNRKLEISSGKLARFADSAVAVQSD 60			
DB	18 DGFLLPRDRALTOQLVRAVAVDIGNRKLEISSGKLARFADSAVAVQSD 77			
QY	61 TAVMTAVSKTPSPSQMPPLVVDYRQKAAAGRIPTNYLREVEGTSKEIITSNIIDS 120			
DB	78 TAVMTAVSKTPSPSQMPPLVVDYRQKAAAGRIPTNYLREVEGTSKEIITSNIIDS 137			
QY	121 IRPLPAPGPFYDTQLGNLAVDGNVNEPDVLAINGASVALSLSDIPLMGPVGAVRIGIID 180			

```

Db 138 IRPLFPAGYFYDTQVLCNLLAVDGNBPDVLAINGASVALSLSDIPMNGPVGAVRIGIID 197
Qy 181 GEYVNPTRKEMSSSTNLVVAAGPKSQIWMLEASAEENILQODPCHAIKGVKXTQOIIQ 240
Db 198 GEYVNPTRKEMSSSTNLVVAAGPKSQIWMLEASAEENILQODPCHAIKGVKXTQOIIQ 257
Qy 241 GIQQLVETGVTKTPOKLTFTSPBEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
Db 258 GIQQLVETGVTKTPOKLTFTSPBEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 317
Qy 301 LDTEBOLKEKPEADPYEIIIESFNVAKEVRSIVLNEYKRCGDRDLSLNVSCEVDMF 360
Db 318 LDTEBOLKEKPEADPYEIIIESFNVAKEVRSIVLNEYKRCGDRDLSLNVSCEVDMF 377
Qy 361 KTLHGSALFQRCQOVCTYTPDSLESGIKSDQVITAINGIKDNFMHLYEPYPATNEI 420
Db 378 KTLHGSALFQRCQOVCTYTPDSLESGIKSDQVITAINGIKDNFMHLYEPYPATNEI 437
Qy 421 GKVTGLNRRELGHGALAEKALYPVLPDPFTIRVTSVLESGSSMASACGSLAMD 480
Db 438 GKVTGLNRRELGHGALAEKALYPVLPDPFTIRVTSVLESGSSMASACGSLAMD 497
Qy 481 SGVPISSAVAGVALGLVTKTDPKEGEIEDYRLTDIIGIEDYNGMDPKIAGTKGITAL 540
Db 498 SGVPISSAVAGVALGLVTKTDPKEGEIEDYRLTDIIGIEDYNGMDPKIAGTKGITAL 557
Qy 541 QADIKLPGIPKIYMEAIQOASVAKKEILOIMNTISKPRASRENGVETVQVPLSKR 600
Db 558 QADIKLPGIPKIYMEAIQOASVAKKEILOIMNTISKPRASRENGVETVQVPLSKR 617
Qy 601 AKFVPGPGYNLKKLQAEETGVTSQVDEETFSVPAPTSVMHEARDFTEICKDQEOOLE 660
Db 618 AKFVPGPGYNLKKLQAEETGVTSQVDEETFSVPAPTSVMHEARDFTEICKDQEOOLE 677
Qy 661 FGAVYTAITEIRDTGVWVKLYPNMTAVLLHNTQLDNERL 700
Db 678 FGAVYTAITEIRDTGVWVKLYPNMTAVLLHNTQLDQSKI 717

RESULT 2
Q8TCS8 PRELIMINARY; PRT; 783 AA.
ID Q8TCS8 AC Q8TCS8
DT 01-JUN-2002 (Tremblrel. 21, Last Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Polynucleotide phosphorylase-like protein (EC 2.7.7.8).
GN PNPase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBT_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rajmikeers R., Vree Egberts W., Van Venrooij W., Pruijn G.;
RT "Protein-protein interactions between human exosome components suggest
RT the assembly of RNase H-type subunits into a six-membered PNPase-like
RT ring."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ458465; CAD30289.1; -.
DR InterPro; IPR001247; 3 Exonase.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR003029; S1.
DR Pfam; PF03726; PNPase; 1.
DR Pfam; PF01138; RNase PH; 2.
DR Pfam; PF03725; RNase PH_C; 2.
DR Pfam; PF00575; S1; 1.
DR SMART; SM00322; KH; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

```

```

DR PROSITE; PS50084; KH_type_1; 1.
DR PROSITE; PS50126; S1; 1.
KW Transferase; Nucleotidyltransferase.
SQ SEQUENCE 783 AA; 85936 MM; 8A3629AF5528E24 CRC64;

Query Match 98.7%; Score 3512; DB 4; Length 783;
Best Local Similarity 99.3%; Pred. No. 5.3e-219;
Matches 695; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DGPELRRPRDALTOQLVRLMSAGSRAVAVDGNKRLTSSGKLRFADGSAVVOGSD 60
Db 18 DGPELRRPRDALTOQLVRLMSAGSRAVAVDGNKRLTSSGKLRFADGSAVVOGSD 77
Qy 61 TAVVTVASKTSPSPQFMPVLVDYRQKAAAGRIPTNYLREYGTSDKEILTSRIIDRS 120
Db 78 TAVVTVASKTSPSPQFMPVLVDYRQKAAAGRIPTNYLREYGTSDKEILTSRIIDRS 137
Qy 121 IRPLFPAGYFYDTQVLCNLLAVDGNBPDVLAINGASVALSLSDIPMNGPVGAVRIGIID 180
Db 138 IRPLFPAGYFYDTQVLCNLLAVDGNBPDVLAINGASVALSLSDIPMNGPVGAVRIGIID 197
Qy 181 GEYVNPTRKEMSSSTNLVVAAGPKSQIWMLEASAEENILQODPCHAIKGVKXTQOIIQ 240
Db 198 GEYVNPTRKEMSSSTNLVVAAGPKSQIWMLEASAEENILQODPCHAIKGVKXTQOIIQ 257
Qy 241 GIQQLVETGVTKTPOKLTFTSPBEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
Db 258 GIQQLVETGVTKTPOKLTFTSPBEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 317
Qy 301 LDTEBOLKEKPEADPYEIIIESFNVAKEVRSIVLNEYKRCGDRDLSLNVSCEVDMF 360
Db 318 LDTEBOLKEKPEADPYEIIIESFNVAKEVRSIVLNEYKRCGDRDLSLNVSCEVDMF 377
Qy 361 KTLHGSALFQRCQOVCTYTPDSLESGIKSDQVITAINGIKDNFMHLYEPYPATNEI 420
Db 378 KTLHGSALFQRCQOVCTYTPDSLESGIKSDQVITAINGIKDNFMHLYEPYPATNEI 437
Qy 421 GKVTGLNRRELGHGALAEKALYPVLPDPFTIRVTSVLESGSSMASACGSLAMD 480
Db 438 GKVTGLNRRELGHGALAEKALYPVLPDPFTIRVTSVLESGSSMASACGSLAMD 497
Qy 481 SGVPISSAVAGVALGLVTKTDPKEGEIEDYRLTDIIGIEDYNGMDPKIAGTKGITAL 540
Db 498 SGVPISSAVAGVALGLVTKTDPKEGEIEDYRLTDIIGIEDYNGMDPKIAGTKGITAL 557
Qy 541 QADIKLPGIPKIYMEAIQOASVAKKEILOIMNTISKPRASRENGVETVQVPLSKR 600
Db 558 QADIKLPGIPKIYMEAIQOASVAKKEILOIMNTISKPRASRENGVETVQVPLSKR 617
Qy 601 AKFVPGPGYNLKKLQAEETGVTSQVDEETFSVPAPTSVMHEARDFTEICKDQEOOLE 660
Db 618 AKFVPGPGYNLKKLQAEETGVTSQVDEETFSVPAPTSVMHEARDFTEICKDQEOOLE 677
Qy 661 FGAVYTAITEIRDTGVWVKLYPNMTAVLLHNTQLDNERL 700
Db 678 FGAVYTAITEIRDTGVWVKLYPNMTAVLLHNTQLDQSKI 717

RESULT 3
Q8KLR3 PRELIMINARY; PRT; 783 AA.
ID Q8KLR3 AC Q8KLR3
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Polynucleotide phosphorylase-like protein (EC 2.7.7.8).
GN 1200003F12RIK OR PNPase.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBT_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN-divers; Vree Egberts W., van Venrooij W., Pruijn G.J.M.;
 RA Rajmakers R., Proteins-protein interactions between human exosome components support
 RT the assembly of RNase PH-type subunits into a six-membered RNase-like
 RT ring";
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AJ507387; CAD45436.1; -;
 DR MGD; MG1:1918951; 1200003F12Rik.
 DR InterPro; IPR001247; 3 ExoRNase.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR InterPro; IPR004087; KH_dom.
 DR InterPro; IPR004088; KH_type_1.
 DR InterPro; IPR003029; S1.
 DR Pfam; PF03726; RNase; 1.
 DR Pfam; PF01138; RNase PH; 2.
 DR Pfam; PF03725; RNase PH_C; 2.
 DR Pfam; PF00575; S1; 1.
 DR SMART; SM00322; KH; 1.
 DR SMART; SM00316; S1; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR PROSITE; PS50084; KH_TYPE_1; 1.
 DR PROSITE; PS50126; S1; 1.
 KM Transferase; Nucleotidyltransferase.
 SQ SEQUENCE 783 AA; 85682 MW; F35F6B591AAB5626 CRC64;

Query Match 92.2%; Score 3280; DB 11; Length 783;
 Best Local Similarity 91.0%; Pred. No. 5 8e-204;
 Matches 636; Conservative 39; Mismatches 24; Indels 0; Gaps 0;

QY 2 GPELPRDRALTOQLVRLMSSAGSAVADLGNKLEISSGKLARFADGSAVVGSGDT 61
 DB 19 GLGRGRRRLALSYLQMRALMSTSGRAVTVLGRHKLKISSGKLARFADGSAVVGSGDT 78
 QY 62 AMMTAVSTKSPSPQFMPLVVDYRQKAAAGRIPIINYLREVGISDKILISRIIDRI 121
 DB 79 AMMTAVSTKSPSPQFMPLVVDYRQKAAAGRIPIINYLREVGISDKILISRIIDRI 138
 QY 122 RPLPAGVGYPTQVLCNLAADGVNPEVLAINGASVALSLDIPNNGVGAVRIGIID 181
 DB 139 RPLPAGVGYPTQVLCNLAADGVNPEVLAINGASVALSLDIPNNGVGAVRIGIID 198
 QY 182 EYVNPTRKEMSSSTLNLVAVGAPKSOIWMLEASAEINLQDFCHAIKVGKYYTQIIIG 241
 DB 199 EYVNPTRKEMSSSTLNLVAVGAPKSOIWMLEASAEINLQDFCHAIKVGKYYTQIIIG 258
 QY 242 IQQLVKEGVTKRTPOKLFTEPSELYKTHKLAERLVAFTDYEDKSRDEAVNKIRL 301
 DB 259 IQQLVKEGVTKRTPOKLFTEPSELYKTHKLAERLVAFTDYEDKSRDEAVNKIRL 318
 QY 302 DPEEOLKEKPEADPEIIESFNVAKEVRSIVLNEYKRCGRDLTSLRNVCEDMFK 361
 DB 319 DPEEOLKEKPEADPEIIESFNVAKEVRSIVLNEYKRCGRDLTSLRNVCEDMFK 378
 QY 362 TLHGSALFORQOTVLCVTFPDSLESIGKSDQVITAINGIKDKNFMLHYEPFPYAT 421
 DB 379 TLHGSALFORQOTVLCVTFPDSLESIGKSDQVITAINGIKDKNFMLHYEPFPYAT 438
 QY 422 KTTGLARRELGHGALAEKALYVIPDPEFTIRVTSSEVLESNGSSSMASACGSLALMDS 481
 DB 439 KTTGLARRELGHGALAEKALYVIPDPEFTIRVTSSEVLESNGSSSMASACGSLALMDS 498
 QY 482 GVPISSAVAGVAGLVTCTDPEKGEIDYRLTLGLIEDVNGMDPFKAGTNKIGITLMO 541
 DB 499 GVPISSAVAGVAGLVTCTDPEKGEIDYRLTLGLIEDVNGMDPFKAGTNKIGITLMO 558
 QY 542 ADIKLPGIPIKIWEALIQASVAKKEILQINNKTIKSPRSRKNKNGPVETVQVPLSKRA 601
 DB 559 ADIKLPGIPIKIWEALIQASVAKKEILQINNKTIKSPRSRKNKNGPVETVQVPLSKRA 618
 QY 602 KVVGGGNYLKKLQAEVGTISQVDEETFSVAFPTPSVMHARDPTTEICKDDOEOLEF 661
 DB 619 KVVGGGNYLKKLQAEVGTISQVDEETFSVAFPTPSVMHARDPTTEICKDDOEOLEF 678

QY 662 GAVYATITTEIRDTGVWVKLYPNMTAVLLHNTOLDNERL 700
 DB 679 GAVYATITTEIRDTGVWVKLYPNMTAVLLHNSQLDQRI 717

RESULT 4

08R2U3 PRELIMINARY; PRT; 589 AA.

ID 08R2U3
 AC 08R2U3;
 DT 01-JUN-2002 (TRENDEL. 21, Created)
 DT 01-JUN-2002 (TRENDEL. 21, Last sequence update)
 DT 01-MAR-2003 (TRENDEL. 23, Last annotation update)
 DE Similar to putative.
 GN 1200003F12Rik.
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC027228; AAH27228.1; -;
 DR MGD; MG1:1918951; 1200003F12Rik.
 DR InterPro; IPR001247; 3 ExoRNase.
 DR InterPro; IPR004087; KH_dom.
 DR InterPro; IPR004088; KH_type_1.
 DR InterPro; IPR003029; S1.
 DR Pfam; PF03726; RNase; 1.
 DR Pfam; PF01138; RNase PH; 1.
 DR Pfam; PF03725; RNase PH_C; 1.
 DR SMART; SM00322; KH; 1.
 DR SMART; SM00316; S1; 1.
 DR PROSITE; PS50084; KH_TYPE_1; 1.
 DR PROSITE; PS50126; S1; 1.
 SQ SEQUENCE 589 AA; 64962 MW; 808162BBA6F6428F CRC64;

Query Match 69.8%; Score 2484; DB 11; Length 589;
 Best Local Similarity 92.2%; Pred. No. 1 4e-152;
 Matches 482; Conservative 26; Mismatches 15; Indels 0; Gaps 0;

QY 178 IIDGEVNPTRKEMSSSTLNLVAVGAPKSOIWMLEASAEINLQDFCHAIKVGKYYTQ 237
 DB 1 MIDGCVNPTREKSSSTLNLVAVGAPKSOIWMLEASAEINLQDFCHAIKVGKYYTQ 60
 QY 238 IIOGIIQQLVKEGVTKRTPOKLFTEPSELYKTHKLAERLVAFTDYEDKSRDEAVN 297
 DB 61 IIOGIIQQLVKEGVTKRTPOKLFTEPSELYKTHKLAERLVAFTDYEDKSRDEAVN 120
 QY 298 KIRLDEEOLKEKPEADPEIIESFNVAKEVRSIVLNEYKRCGRDLTSLRNVCEDV 357
 DB 121 KIRLDEEOLKEKPEADPEIIESFNVAKEVRSIVLNEYKRCGRDLTSLRNVCEDV 180
 QY 358 DMFKTLHGSALFORQOTVLCVTFPDSLESIGKSDQVITAINGIKDKNFMLHYEPFPYAT 417
 DB 181 DMFKTLHGSALFORQOTVLCVTFPDSLESIGKSDQVITAINGIKDKNFMLHYEPFPYAT 240
 QY 418 NEIGVTLARRELGHGALAEKALYVIPDPEFTIRVTSSEVLESNGSSSMASACGSLA 477
 DB 241 NEIGVTLARRELGHGALAEKALYVIPDPEFTIRVTSSEVLESNGSSSMASACGSLA 300
 QY 478 LMDGVPISSAVAGVAGLVTCTDPEKGEIDYRLTLGLIEDVNGMDPFKAGTNKIGITLMO 537
 DB 301 LMDGVPISSAVAGVAGLVTCTDPEKGEIDYRLTLGLIEDVNGMDPFKAGTNKIGITLMO 360
 QY 538 TALQADIKLPGIPIKIWEALIQASVAKKEILQINNKTIKSPRSRKNKNGPVETVQVPLSKRA 597
 DB 361 TALQADIKLPGIPIKIWEALIQASVAKKEILQINNKTIKSPRSRKNKNGPVETVQVPLSKRA 420
 QY 598 SKRAKVFPGGNYLKKLQAEVGTISQVDEETFSVAFPTPSVMHARDPTTEICKDDOEO 657
 DB 421 SKRAKVFPGGNYLKKLQAEVGTISQVDEETFSVAFPTPSVMHARDPTTEICKDDOEO 480

Qy 658 QLEFGAVYATITETIRDTGVWVKLYPNMTAVLNLNTOLDNERL 700
 Db 481 QLEFGAVYATITETIRDTGVWVKLYPNMTAVLNLNTOLDNERL 523

RESULT 5

Q9DC52 PRELIMINARY; PRT; 540 AA.
 AC Q9DC52;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE 1200003F12RIK protein.
 GN 1200003F12RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=C57BL/6J; TISSUE=Lung;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai U., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi U., Fukuda S.,
 RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H. A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuehl P., Lewis S., Maeno Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schiraldi L. M., Steubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M. F.,
 RA Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guernicich S., Hill D., Hofmann M., Hume D. A., Kamita M., Lee N. H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Nombere P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
 RA Suzuki H., Toyooka K., Wang K. H., Weitz C., Wiltaker C., Wilmink L.,
 RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlschki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:665-690(2001).
 DR EMBL; AK004563; BAB3374.1; -
 DR MGI; MGI:1918951; 1200003F12RIK.
 DR InterPro; IPR001247; 3 ExonNase.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF03726; ENPase; 1.
 DR Pfam; PF01138; RNase_P; 2.
 DR Pfam; PF03725; RNase_P; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 SQ SEQUENCE 540 AA; 58938 MW; 457BFA3B3579A072 CRC64;

Query Match 67.2%; Score 2389; DB 11; Length 540;
 Best Local Similarity 90.1%; Pred. No. 1.8e-14;
 Matches 465; Conservative 29; Mismatches 22; Indels 0; Gaps 0;

Qy 2 GPELPRDRALTOLOVRALMSAGRAVAVDIGNRKLKISSGKLARFADGSAVVOGSDT 61
 Db 19 GPLRPGNRALSTYLOMRALMSSTGSAVAVDIGNRKLKISSGKLARFADGSAVVOGSDT 78
 Qy 62 AVMTAVSKTSPSPQFMPVLVDYRQKAAAGRIPTNYLRREVSTDEKELTSRIIDRSI 121
 Db 79 AVMTAVSKTSPSPQFMPVLVDYRQKAAAGRIPTNYLRREVSTDEKELTSRIIDRSI 138
 Qy 122 RPLFPAGFYDTQVLCNLLAVDGNPDVLAINGASVALSLSDIPMNGPVGAVRIGIID 181
 Db 139 RPLFPAGFYDTQVLCNLLAVDGNPDVLAINGASVALSLSDIPMNGPVGAVRIGIID 198
 Qy 182 EYVNPTRKEMSSSTLNLVVAAGPKSQIYWLKASANIILQODFCHAIKGVKXTQOIIIG 241
 Db 199 ECVNPTREKSSSTLNLVVAAGPKSQIYWLKASANIILQODFCHAIKGVKXTQOIIIG 258
 Qy 242 IQGLVKGTVKTRPQKLFTSPFBIYKTKHKLAMRLYAVFTDHDKVSDEAVNKIRL 301

Db 259 IQGLVKEIGVAKRTPQKLFTSPFBIYKTKHKLAMRLYAVFTDHDKVSDEAVNKIRL 318
 Qy 302 DTEEQLEKPEPADPYRIISFNVAKEVRSIYLNKRCDDGLTSLNVSCEVMPFK 361
 Db 319 DTEEQLEKPEPADPYRIISFNVAKEVRSIYLNKRCDDGLTSLNVSCEVMPFK 378
 Qy 362 TLHGSALEFGQVQVLCVTFPDSLSGSKSDQVITTAINGIKDNFMLHYEPFYATNEIG 421
 Db 379 TLHGSALEFGQVQVLCVTFPDSLSGSKSDQVITTAINGIKDNFMLHYEPFYATNEIG 438
 Qy 422 KVTGLNRELGKALAKALYPTVPPTIRTSVLSNGSSNMAACGSLAMD 481
 Db 439 KVTGLNRELGKALAKALYPTVPPTIRTSVLSNGSSNMAACGSLAMD 498
 Qy 482 GVPISAVAGVAGLTKTDPKGEIRDYRLITDIL 517
 Db 499 GVPISAVAGVAGLTKTDPKGEIRDYRLITDIL 534

RESULT 6

Q96T05 PRELIMINARY; PRT; 504 AA.
 AC Q96T05;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical protein FLJ14531.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OC NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N. A.
 RA Itoigai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
 RA Watanabe S., Hosoi T., Kaku Y., Kodaira H., Kondo S., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Maeno Y.,
 RA Ninomiya K., Iwayanagi T.,
 RT "NBD human cDNA sequencing project."
 RL Submitted (May-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AK027437; BAB55109.1; -
 DR InterPro; IPR001247; 3 ExonNase.
 DR InterPro; IPR001547; Glyco_hydro_5.
 DR Pfam; PF03726; ENPase; 1.
 DR Pfam; PF01138; RNase_P; 2.
 DR Pfam; PF03725; RNase_P; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 504 AA; 55998 MW; 2BB89AD8409322D6 CRC64;

Query Match 64.2%; Score 2285; DB 4; Length 504;
 Best Local Similarity 100.0%; Pred. No. 8.9e-140;
 Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGPELPRDRALTOLOVRALMSAGRAVAVDIGNRKLKISSGKLARFADGSAVVOGSD 60
 Db 18 DGPELPRDRALTOLOVRALMSAGRAVAVDIGNRKLKISSGKLARFADGSAVVOGSD 77
 Qy 61 TAVMTAVSKTSPSPQFMPVLVDYRQKAAAGRIPTNYLRREVSTDEKELTSRIIDRS 120
 Db 78 TAVMTAVSKTSPSPQFMPVLVDYRQKAAAGRIPTNYLRREVSTDEKELTSRIIDRS 137
 Qy 121 IRPLFPAGFYDTQVLCNLLAVDGNPDVLAINGASVALSLSDIPMNGPVGAVRIGIID 180
 Db 138 IRPLFPAGFYDTQVLCNLLAVDGNPDVLAINGASVALSLSDIPMNGPVGAVRIGIID 197
 Qy 181 GEYVNPTRKEMSSSTLNLVVAAGPKSQIYWLKASANIILQODFCHAIKGVKXTQOIIIG 240
 Db 198 GEYVNPTRKEMSSSTLNLVVAAGPKSQIYWLKASANIILQODFCHAIKGVKXTQOIIIG 257

QY 241 G1Q0LVKGTGKTRTPQKLFPSPEIIVKXTHKLMERLYAVFTDYHDKVSDBAVNKIR 300
 DB 258 G1Q0LVKGTGKTRTPQKLFPSPEIIVKXTHKLMERLYAVFTDYHDKVSDBAVNKIR 317
 QY 301 LDTESQLEKKEPFAPEYIEISFNVAKVPFISVLANEYKRCQDGRDUTSLRVNSCEVDNF 360
 DB 318 LDTESQLEKKEPFAPEYIEISFNVAKVPFISVLANEYKRCQDGRDUTSLRVNSCEVDNF 377
 QY 361 KTLHGSALFQRCQOTVLCVTFPSLESGIKSDQVITANGIKDKKFMFLYEPFPAATNBI 420
 DB 378 KTLHGSALFQRCQOTVLCVTFPSLESGIKSDQVITANGIKDKKFMFLYEPFPAATNBI 437
 QY 421 GAVTGINRRELGHGALAEKALYVPIPRDP 450
 DB 438 GAVTGINRRELGHGALAEKALYVPIPRDP 467
 RESULT 7
 ID Q9V9X7 PRELIMINARY; PRT; 771 AA.
 AC Q9V9X7;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE CG11337 protein.
 GN CG11337.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RX [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Suton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abiri J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Danke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glaeser K.,
 RA Glodek A., Gong P., Correll J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M.B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasco P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milhina N.V., Moadary C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Neufeld D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sideri-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spletter B., Spredling A.C., Stapleton M., Strong R., Sun B.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.M., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).

RN [2]
 RP SEQUENCE FROM N.A.
 RA Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,
 RA Carlson J.M., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Friese R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Matei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Friese B., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,
 RA Seale S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Flybase;
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AB003778; AAF57151.2; -
 DR HSSP; P05055; 1SR0.
 DR Flybase; FBgn0039846; CG11337.
 DR InterPro; IPR001247; 3_E0XRNase.
 DR InterPro; IPR004088; KH_Type_1.
 DR InterPro; IPR003029; S1.
 DR Pfam; PF03726; RNase; 1.
 DR Pfam; PF01138; RNase; 1.
 DR Pfam; PF03725; RNase; PH; 2.
 DR PROSITE; PSS0084; KH_Type_1; 1.
 DR PROSITE; PSS0126; S1; 1.
 DR DR SEQUENCE 771 AA; 84953 MW; 779F086DB17022C4 CRC64;
 SQ
 Query Match 55.0%; Score 1958; DB 5; Length 771;
 Beet Local Similarity 55.2%; Pred. No. 2.7e-118; Indels 8; Gaps 5;
 Matches 380; Conservative 128; Mismatches 173;
 QY 19 PALWSSAGSRAVAVDL--GNRKLKISSGKLARFADGSAVVOGDTAVMTAVSKTKPSP 75
 DB 29 RGIQSSNGEARPSVBNVNSNRNMTFFSGRLARFANGRAVQMDTAVMTAVAKAKENP 88
 QY 76 SQ-FMPLVVDYRQRAAAGRIPTVYLRRVEGTSDEKILITSSIRPLFPAGYFYDTQ 134
 DB 89 GQGFPLVVDYRLKRAASGRIPMNMFRRELGPSEKILISARLIDRSRLPLPHKYRTETQ 148
 QY 135 VLQNLAVDGNVEDPVLAINGASVALISDI.PMNGVQAVAGIIGDGVVNPTRKEMS 194
 DB 149 LVQNLAVDAVHSPVLAINGASVALISDI.PMNGVQAVAGIIGDGVVNPTRKELQT 208
 QY 195 STNLNVVAGAPKSOIVMLEASENILODFCHAIKVGKVKYQOIIQIGIQLVKETGVTKR 254
 DB 209 SQDLVVSATQONVLMLEKGNVLODDLKAIKQGRREAFIHEIRLOKAVGRKXR 268
 QY 255 TPQKLFPSPEIIVKXTHKLMERLYAVFTDYHDKVSDBAVNKIRLDTESQLEKKEPFA 314
 DB 269 EVEVAAEVDPDLGRAVNSMCMERLREIFQDSTHDKMSRDNAVNEVRSNVIDKVVSSFPDT 328

```

OY 315 PVEYELIESPNVAKVERISIVLNEYKRCOGRODITSLRANVSCEYDEMKETHGSLFQRCOT 374
Db 339 EBSLITBOCNGQSRKIFRELIFERGRLCCGRDDQJRNISICQYDWMKPLHGSLFQRCOT 388
OY 375 QVLTCTVPEDLSSEGIKSDOVITAIN--GIKDKNFMLHYEPPPYATIEIGKVTGLNRELG 432
Db 389 QVCTVSLSDSGSAMLKDS-LMALDSGGLKAKNFMLHYEPPPYATIEIGVGRREMG 447
OY 433 HGALAEKALYPVIRPRDPEPTIRATSEVLSNGSSSNASACGSLALMDSGVPTISSAVAGV 492
Db 448 HGALAEKRLPLPLPNDPYPTVRLTSEVLSNGSSSNASACGSLALMDSGVPSAPAGV 507
OY 493 AIGVATK-TDPEKGELEDYRLITDLIGIEDYNDMDMFKIAGTKNGKGTALQADIKLPIPI 551
Db 508 AIGVATKKEENDDTKHLQDYRLITDLIGIEDYNDMDMKAAGTKGFALQADIKLPIPL 557
OY 552 KIWEBAIQOASVAKKEIILQIMNKTISKPRASRENGPVEVTVGPIKRAFVGPGEVNL 611
Db 568 KYMVESLOKATKANSIIDLIMSEALIEPKPKPRESHPVSETTLVEBPQAROLIGPGLHM 627
OY 612 KKLQAEITGVLTISQVDBETFSVPAFTPSVMEHARDFITEICKDQEOOLEGAVYATITE 671
Db 628 KRIYLETGSLTAVIDETHFNVPAPSOAMDDEAKELIEGVVWKERVPDLBERGIGIYTAKITE 687
OY 672 IROTGVMVYGLPMTATVALLHNTQOLDNERL 700
Db 688 LBDTGVMVYLYPSMPALHNSLODRKI 716

```

```

RESULT 8
095RX7
ID 095RX7 PRELIMINARY; PRT; 720 AA.
AC 095RX7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
DE LDOJ255D.
GN CG11337.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachytera; Muscomorpha;
CC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_taxid=7227;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Berkely;
RA Stapleton M., Chavez C., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Birkstein P., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guatin H., Li P., Liao G., Miranda A., Mungall C.U.,
RA Nunoo J., Pacble J., Paragas V., Park S., Phouneavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Ceiniker S.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY061061; AAL28609.1; -.
DR FlyBase; FBgn0039846; CG11337.
DR InterPro; IPR001247; 3_ExonNase.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR003029; S1.
DR Pfam; PF03726; PNase; 1.
DR Pfam; PF01139; RNase_PH_C; 2.
DR Pfam; PF03725; RNase_PH_C; 2.
DR SMART; SM00322; KH; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PSS0084; KH_TYPE_1; 1.
DR PROSITE; PSS0126; S1; 1.
SQ SEQUENCE 720 AA; 79382 MW; DD549525E5BDF CRC64;
Query Match 54.6%; Score 1940.5; DB 5; Length 720;
Best Local Similarity 56.2%; Pred. No. 3,3e-117;
Matches 374; Conservative 123; Mismatches 164; Indels 5; Gaps 4;

```

```

Db 1 MFTSSGRLARFANGTAVCMGBTAPMWTAVAKAKPNPCQGMPLVDVYRLKNAASG1PM 60
Oy 98 NYLRREVTSDEKELTSRIIDRSIRPLFPAGFYDOTOVLCNLAVDGNBPDVLAINGAS 157
Db 61 NFMREIRGESEKEILLBARLIDRSPLRPHKORYRTETOVLVCMMLAMDVAHSPDVLAIWAAS 120
Oy 158 VALSLSDIPMNGVGVAVRIGIIDGSEYVNPTRKEMSSSTLNLVAAGA PKSDIWLJASAE 21.7
Db 121 MALSLSDIPMNGPIGAVRVGLCDSEVLNPFTRRELQTSQDLVLSATKONLVMLBEGKN 180
Oy 218 NILQODFCHAIRVGVKYTOOIIQGIQOLVKETGVKTRPOKLFTPSBEIYKTHKLAMER 277
Db 181 VVLQODLLKAIKQGTREAOPIIHEIRLQKAVGRQKREVEVAABVDDELCAVSMCEMR 240
Oy 278 LVAVPFYDHDKYSRBEAVNKRILDTBEOLEKEPEADPYEIIISPNVNAKEVPRSTVLN 33.7
Db 241 LREIPFDSTHDKSRDANAVENRNVNIDKWSSPFDTBESLITQFQNTSKTIFRELIFE 300
Oy 338 EYRCDCGRDLTSLRNVCCEVDMFKTLHGSALEFORQOTOVLCTVFPDSLESGIKSDQVITA 39.7
Db 301 RGRKCDGRDYDOLRNISQGVDMYKPLHGSALFQSGQTOVCTVSLDSEGSAMKUDS-LAA 3.59
Oy 398 IN-GIKOKNFMHLHEPPYATNEIGKYOTGLNRRELHGALAEKALYPVIRPPDPFITRV 4.55
Db 360 LBDGGKAKAKNFMHLHEPPYATGEBVGRIGPVRGRMRMGALAEBSLLPTLPNDVPFVRL 4.19
Oy 456 TSEVLESNGSSSSMAASCGGSLMLMDSGVPISAVNGVAILGVTK-TDPEKEIIDEYLLT 51.4
Db 420 TSEVLESNGSSSSMAASCGGSLMLMAGVPVSPAAGVAILGVTTFFENDORTHLDOYIILT 4.79
Oy 515 DLIGIEDYNGMDXFIAGTNKGITALLQADIKLPGIPIKIWEALIQOASVAKKEILLQIMNK 57.4
Db 480 DILIGIEDYNGMDXMKVAGTRKGFALIQADLKIPIGLKVWESLQKADAKSNLIDIMSE 53.99
Oy 575 TIEKPRASKEKNGPVETVOYPLSRARFVPCGGINLKKLOAFNGCVTISOUDETBSEVFA 63.4
Db 540 AIEBPKYIPESBPVSEITLTVPEQPORAOIIIGSGIHMKRILYLEGTSLTADDEHFNVFA 59.99
Oy 635 PTFSVWHEARDFTTEICKDDEQOQEFAGAVYATITTEIRDGNWVKKYPMNWTAVLLANTQ 69.4
Db 600 PQGAAMDAXEKELLEGMVMEKVRPDLPEFGIYTAKTITELRDTGVWVILLYPSMPPALLNSQ 65.99
Oy 695 LDNERL 700
Db 660 LDQRKI 665

```

RESULT	9			
Q8IH29		PRELIMINARY;	PRT;	717 AA.
AC	Q8IH29			
DT	01-MAR-2003	(T-EMBLrel. 23, Created)		
DT	01-MAR-2003	(T-EMBLrel. 23, Last sequence update)		
DT	01-MAR-2003	(T-EMBLrel. 23, Last annotation update)		
DE	GM6802P (Fragment).			
GN	CG11337.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephyridiidae; Drosophilidae; Drosophila.			
OX	NCBI_TaxId=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,			
RA	Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,			
RA	Gerardo R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,			
RA	Miranda A., Mungall C.J., Nunco J., Pachle J., Paragas V., Park S.,			
RA	Patel S., Phouenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,			
RA	Celniker S.;			
RL	Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BT001457; AAN71212.1; --.			
ET	NON TER			
SO	SEQUENCE			
	717 AA; 79404 MW; 6C1CA7ABCI714D02 CRC64;			

Query Match 49.5%; Score 1761; DB 5; Length 717;
 Best Local Similarity 54.4%; Pred. No. 1.4e-105;
 Matches 336; Conservative 119; Mismatches 159; Indels 4; Gaps 3;

QY 86 ROKAAAGRIPIRYARREVGTSKILTSRIIDRSIRPLFPAGFYDTQVLCNLAVDGV 145
 DB 46 RMTSSGRIPIRYARREVGTSKILTSRIIDRSIRPLFPAGFYDTQVLCNLAVDGV 105
 QY 146 NEPDVLAINGASVALSLSDIPMNGPVGARIGIIGEEYVNPTRKMSSTLNLVAGAP 205
 DB 106 HSPDLVLAINGASVALSLSDIPMNGPVGARIGIIGEEYVNPTRKMSSTLNLVAGAP 165
 QY 206 KSOIYVLEASAEIILQDFCHAIKGVKVTQOIIQGIQOLVETGVTYKTPKLTFTSPSE 265
 DB 166 QNLVYVLESGKNVVLQDILKAIKQGTREAOFTIHIEIRLQKAYGQKREVAVAEVDPE 225
 QY 266 IYKTHKLAMERLYAVFTDYEDKYSRDEAVNKIRLDTBEOLKEKPEADPYETIISFNV 325
 DB 226 LKAVRSMCEMLRRIEIPDSTHDKSRDAVNEVRSNVLDKWSFPPDTEPDLITEQFNO 285
 QY 326 VAKVEFRSLVINEYKRCDDGRLTSLRNVCEDVMEFTLHGSALFORGOTVCTVTPSL 385
 DB 286 TSRTIFREILFERGRCDDGRLTSLRNVCEDVMEFTLHGSALFORGOTVCTVTPSL 345
 QY 386 ESGIKSDOYITAIN--GIKDKNFMHYEPFPATNIEIGVTGLNREILGHALAKALYP 443
 DB 346 ESAMKLDS-LAALDSGGLAKKPMHYEPFPATNIEIGVTGLNREILGHALAKALYP 404
 QY 444 VLRPPPTIRVTSYLEVSSNGSSSMASACGSLALMDGVPISSAVAGVAILVTK-TDP 502
 DB 405 TLPNVYFPVRLTSEVLEBSNGSSSMASACGSLALMDGVPISSAVAGVAILVTKFEND 464
 QY 503 EKGIEDYRLTDIIGIEDYNGDMFKIAGTKGITALQADIKLGIPKIYMEALQAS 562
 DB 562 DTKHLDYRLTDIIGIEDYNGDMFKIAGTKGITALQADIKLGIPKIYMEALQAS 524
 QY 524 VAKKEILQIMNTKISPRASRKENGVPVETVPLSKRAKFPVPGGVNKLQAEYVTI 622
 DB 622 DAKSILIDMSAIRKPKESWPSFTLVBEVQKQALIGPSGLHAKRIYLETGSL 584
 QY 622 SOVDEETSVPAPTSVMEHARDFTTEICKDOQOEGLFAGVYVTAITEIRDTGVWVLY 682
 DB 682 TAVDETFVFPAPTSVMEHARDFTTEICKDOQOEGLFAGVYVTAITEIRDTGVWVLY 644
 QY 683 PMNTAVLHNTOLDNERL 700
 DB 645 PSMPPALLHNSQLDQRI 662

RESULT 10
 Q9S7G6 PRELIMINARY; PRT; 991 AA.
 AC Q9S7G6; 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
 DE Polynucleotide phosphorylase.
 GN PNP OR T15N1_70.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsia.
 CC NCBI_TaxId=3702;
 CX [1]
 RN SEQUENCE FROM N.A.
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Newes H. W.,
 RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
 RD [2]
 RA Rudd S., Lemcke K., Mayer K.F.X.;

RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL, Y14686; CAB43865.1; -
 DR EMBL, Y14685; CAB43864.1; -
 DR EMBL, AL163792; CAB87625.1; -
 DR HSPD; P05055; ISRO.
 DR InterPro; IPR001247; 3 Exonase.
 DR InterPro; IPR004087; K1_dom.
 DR InterPro; IPR003029; S1.
 DR Pfam; PF03726; PNPase; 1.
 DR Pfam; PF01138; RNase PH; 2.
 DR Pfam; PF00575; S1; 1_C; 2.
 DR SMART; SM00322; KH; 1.
 DR SMART; SM00316; S1; 2.
 DR PROSITE; PS50126; S1; 1.
 KW Transic peptide.
 SQ SEQUENCE 991 AA; 107771 MW; 9557E2980C6D1AS CRC64;

Query Match 38.4%; Score 1365.5; DB 10; Length 991;
 Best Local Similarity 41.7%; Pred. No. 1e-79;
 Matches 287; Conservative 138; Mismatches 231; Indels 33; Gaps 11;

QY 24 SAGSRV-----AVDAGRKLEISGKLFAPDGSAAVQSGDTAVMTAVSKTKSPQF 78
 DB 44 SAGTKILSFEERFVSGRVVSFTGKLFARANGSVAGMDETKVSLTVTCAKTDSPDF 103
 QY 79 MVLVDYRQKAAAGRIPTNYLRRVGTSDKEILTSRIIDRSIRPLFPAGFYDTQVLCN 138
 DB 104 LPLVTDYQKQYAGLINTYMRREGAPKRELLCGRLIDRPLFTGTFHEVQIMAS 163
 QY 139 LLAADVNEPVLAINASVALSLSDIPMNGPVGARIGIIGEEYVNPTRKMSSTLN 198
 DB 164 VLSSGKODPDILANASSALMLSDVPGGPIGIRIGRIGCGFVNPVPMDELSDDL 223
 QY 199 LVVGAPEKQVLMASAEIILQDFCHAIKGVKVTQOIIQGIQOLVETGVTYKTPKLTFTSPSE 254
 DB 224 LIVA-CTNDKTMMDIVQREISEKDLAALALHPEAKVYLDPOI---RLAEKAKQK 278
 QY 255 TPQKLTSPPEIVYKTHKLAMERLYAVFTDYEDKYSRDEAVNKIRLDTBEOLKEKPEA 314
 DB 279 -EYKLSMISDKTLEKVDLALATRIESVFTDPSYGFEEGALDNIGKQVRYFEEGQOE 337
 QY 315 DPEIIESFNVAKEFRSIVLANEYKRCDDGRLTSLRNVCEDVMEFTLHGSALFORGOT 374
 DB 338 SLSTLPKAVDTVRKRVSRMISDGFVYDGHVDEVRPIYCESHYLPALHGSALFSPGDT 397
 QY 375 QVLCTVTFDSLESGIKSDOYITAINGIIDKPMHYEPFPATNIEIGVTGLNREILGH 434
 DB 398 QVLCTVTFDSLESGIKSDOYITAINGIIDKPMHYEPFPATNIEIGVTGLNREILGH 453
 QY 435 ALAEKALYVIRPD--PFTIRVTSYLEVSSNGSSSMASACGSLALMDGVPISSAVAGV 492
 DB 454 TLAEKALYVIRPD--PFTIRVTSYLEVSSNGSSSMASACGSLALMDGVPISSAVAGV 513
 QY 493 AIGLVTKDPEKIEDYRLTDIIGIEDYNGDMFKIAGTKGITALQADIKLGIPKIY 552
 DB 514 SVGLITVDVPSGEEKDRIYVTDILGLEDHGDMDFKIAGTRDGTALQIDIKPGLD 573
 QY 553 IYMEALQOASVAKKEIILQIMNTKISPRASRKENGVPVETVPLSKRAKFPVPGGVNKL 612
 DB 574 IVCSLENAERARLQIDHMERINNSPGQGAISPRATTKYSNLSRLTILGPGVLR 633
 QY 613 KLOAETGVTLISQVDETSVPAPTSVMEHARDFTTEICKDOQOEGLFAGVYVTAITE 669
 DB 634 KIVETGARLS-LINGTLITIAKQDWEKAKQEOVDFTI-----GRELVAVGIVKGTV 685
 QY 670 TEIRDTGVWVLYPMNTAVLHNTOLDNER 698
 DB 686 SSIKEYGAFVE-FPGQGGGLHMSLSHE 713

QY 267 VYTHKLANE-----RLVAVFTDY-----HDKVSDEAVNKRILDTBEOLKEKPE--AD 315
 Db 233 V--LHEDIDLEBEKVAAYATEKYNALRTPEKKERNDNDKVEQVLEHFKDEYEDNLAD 290
 QY 316 PEEIESFVNAKEVRSIVLNEYKCDGRODLSLANVCEVDMFTLHGSALFORGOTO 375
 Db 291 IDEVLYK---INKEQNRKMIKEEKIRVDGRGLDDIRPICEVGVLPRTGSAIFRTGOTO 347
 QY 376 VLCTVTPDLSBSGKSDOVITAINGIKDNFMHYEPFYATNEIGKVTGLNRRELGHGA 435
 Db 348 VLVATL-----GALIGIQLLEGIDGDEFKRYMHMHPYISGEVAPLPGRGRELGHGA 403
 QY 436 LAEKALYPVP--RDPFTIRVTEVLESNGSSSMASACGSLALMDSGVPISSAVAGA 493
 Db 404 LERLLEPIVPSSEEPYITRLVSEVLSNGSTSOASVCGSTLALMDAGVPIKAPVAGIA 463
 QY 494 IGLVTKTDEKGEIEDYRLTLIDIGIEDYNGDMDFIAGTNGKITLQADIKLPGIPIKI 553
 Db 464 MGLIIEED-----EVLLITDIOGIEDEPLGDMDFVAGTEKGVTAIQMDIKIPGIDRDI 516
 QY 554 VNEATQOASVAKKEILQINNKITSKRBSARKENGVPVETVOVPLSRKAFVGVGKYNLKK 613
 Db 517 LQMLAEKAKARLYVQKMLEVIEKEPRKELSKYAPRVAVMVPKIRDIIGPAGKTIYK 576
 QY 614 LOAETGVITISQVDEETFSVPAPTPSVMEHARDPITEICDDOEQOLEFGAVYATITTEIR 673
 Db 577 IISGVKIDIEDGRLYITAPLSEGERAKOMIEAITD-----IEVGITLGVKYLRLA 631
 QY 674 DTGVAVKLYPNMTAVILANTQDNERL 700
 Db 632 PGAFVETIAPGRKG--LVHISNLSKKRV 657

RESULT 13

ID 097145 PRELIMINARY; PRT; 703 AA.
 AC 097145;
 DT 01-OCT-2001 (TREMblrel. 18, Created)
 DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 Db POLYribonucleotide nucleotidyltransferase.
 GN CAC1808.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 CX NCBI_Taxid=1488;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Gibbons J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
 Tatusov R.L., Lee H.M., Dubois J., Qiu D., Hitt J., Wolf Y.I.,
 Benett G.N., Koonin E.V., Smith D.R.,
 RA "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum.";
 RL J. Bacteriol. 183:4823-4838(2001).
 DR EMBL; AB007689; AAK79773.1; -
 DR InterPro; IPR001247; 3_ExonNase.
 DR InterPro; IPR004087; KH dom.
 DR InterPro; IPR004088; KH_type_1.
 DR InterPro; IPR003029; SI.
 DR Pfam; PF00013; KH.1.
 DR Pfam; PF03726; PNPase.1.
 DR Pfam; PF01138; RNase_PH.2.
 DR Pfam; PF03725; RNase_PH_C.2.
 DR Pfam; PF00575; SI.1.
 DR SMART; SM00322; KH.1.
 DR SMART; SM00316; SI.1.
 DR PROSITE; PS50084; KH_type_1.1.
 DR PROSITE; PS50126; SI.1.
 KW transferase; Complete proteome.

SO SEQUENCE 703 AA; 77989 MW; 5EABCB8EDEDCA22B CRC64;

Query Match 33.9%; Score 1204.5; DB 16; Length 703;

Best Local Similarity 39.6%; Pred. No. 1,66-69;

Matches 266; Conservative 125; Mismatches 258; Indels 23; Gaps 10;

QY 33 DLGNKLEISSGKLARFADGAVVOSGDTAVVAVTSKTSPPS-QFMPLVVDYRQKAA 91
 Db 8 DIAGKRLKVECGKTGMLNCAFMFISYGDYVAVVAVNAEKEPREGIDFFPLSIEYERQVS 67
 QY 92 AGRITVTLRREVGSDSEILTSRIIDRSIRPLPAGFYPTQVLCNLANDVNEPDL 151
 Db 68 VKIIPGVKREGRSEKSIILHARAIIDRLPLFPKGRANDVQVCTWASVEQDMLPEIL 127
 QY 152 AINGASVLSLSDIPMNPGVAVRIGIIDGKVVPPTREKSSSTLNVVAGAPSOIYM 211
 Db 128 ANNGASMLCLSDIPFTTVPATVSVCIDGKFLVPLTBEKSSSLDITVC-ATNERVWM 186
 QY 212 LEASANTLODPFCHAIKGVAVYTOQIIQGIQOLVKETGVTKRTPQKFTPSPEIVKYTH 271
 Db 167 LEAGDAEIPEDLMIAIDFGFNAQDDIVAFQEKAMKEGKEKVTPE-LYHPKEIEKQVT 245
 QY 272 KLANERLYAVFTDVBHDVSRDEAVNKRILDTBEOLKEKPEPADYEIIESFVNAKEVF 331
 Db 246 EFAPESIKEIM--YITDDEBNRLRLREIKEXISNEFAEKYPD-DEADIDEVVYTLQKKV 302
 QY 332 RSVINEXKRCQDGRDLTSLRVNVCEDVMFKTLHGSALFORGOTOVLCTVTPDLSBSGKS 391
 Db 303 RNMLKEHRRPDGRFDEIRPLSCVDLPTHTSGGLTRGLTQVWYVYTL---GPIDG 358
 QY 392 DOVITAINGIKDNKFMHYEPFYATNEIGKVTGLNRRELGHGALEKALYPVP--RDF 449
 Db 359 AQVIDGLVESKRYMHMHPYISGEVAPLPGRGRELGHGALEKALYPVP--RDF 418
 QY 450 PFTIVTESEVLESNGSSSMASACGSLALMDSGVPISSAVAGVAGLYTKTDPKEGIED 509
 Db 419 PYTRLVSEVLSNGSTSOASVCGSTLALMDAGVPIKRAPAGIAGLITSEBLSKEAV-- 476
 QY 510 YRLTLIDIGIDVNDMPFKIAGTNGKITLQADIKLPGIPIKIYMEALQOASVAKKEIL 569
 Db 477 ---ITDIOGLSDPFEDMDPFVAGTEKGYITAIQVDTKHLSKYCKITAINDRKARLFTL 533
 QY 570 QIMNKITSKRBSARKENGVPVETVOVPLSRKAFVGVGKYNLKKLOAETGVITISQVDEET 629
 Db 534 EKMVACINERPKELSTVAPRAYTINIDDKIRTLIGTGKTIINKIIEBTGVKIDIREGT 593
 QY 630 FSVFAPTPSVMEHARDPITEICDDOEQOLEFGAVYATITTEIRDTGVAVKLYPNMTAVL 689
 Db 594 VFVLSDDASNRALKMIDLTLD-----VKGEVYLGKVTITFGAFVEVLPKEG-L 647
 QY 690 LANTQDNERLN 701
 Db 648 WHISKLDINKVN 659

RESULT 14

ID 08KBY3 PRELIMINARY; PRT; 733 AA.
 AC 08KBY3;
 DT 01-OCT-2002 (TREMblrel. 22, Created)
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 Db POLYribonucleotide nucleotidyltransferase.
 GN PNP OR CT1649.
 OS Chlorobium tepidum.
 OC Bacteria; Chlorobi; Chlorobiales; Chlorobiaceae;
 OC Chlorobium.
 CX NCBI_Taxid=1097;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 49652 / DSM 12025;
 RX MEDLINE=22103685; PubMed=12093901;
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,

RA Dodson R.J., Debroy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA Vanatrevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA Venter J.C., Tettelin H., Bryant D.A., Frazer C.M.,
RT "The complete genome sequence of *Chlorobium tepidum* TLS, a
RT photoautotrophic, anaerobic, green-sulfur bacterium,"
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
DR EMBL: AE012920; AAM72874.1; -.
DR TIGR: CT1649; -.
DR InterPro: IPR001247; 3_ExoRNase.
DR InterPro: IPR004087; KH_dom.
DR InterPro: IPR004088; KH_type_1.
DR InterPro: IPR003029; SL.
DR Pfam: PF00013; KH; 1.
DR Pfam: PF03726; PNPase; 1.
DR Pfam: PF01138; RNase_PH; 2.
DR Pfam: PF03725; RNase_PH_C; 2.
DR Pfam: PF00575; SL; 1.
DR SMART: SM00322; KH; 1.
DR SMART: SM00316; SL; 1.
DR PROSITE: PS50084; KH_TYPE_1; 1.
DR PROSITE: PS50126; SL; 1.
DR Transferrase; Complete proteome.
SQ SEQUENCE 733 AA; 79937 MW; 3C5D0C71E00C6AD9 CRC64;

Query March 33.7%; Score 1198; DB 16; Length 733;
Best Local Similarity 38.1%; Pred. No. 4.6e-69;
Matches 266; Conservative 144; Mismatches 233; Indels 56; Gaps 16;

32 VDLGNRK-LEISSGKLAFADGSANVOSGDTAVMTAV-SKTKPSPG-FMPLVVDYRQK 88
7 IDLGHGKVIETIGGMAKQADGSANVTMTAVTAVSSKTPSPNDPFLQVEYREK 66
89 AAAAGRIPTNVLREVGTSDEKELTSRIIDRSIRPLPAGFYDTQVLCNLAADVNER 148
67 YSAAGKPGFGFKRGRSEKEILSARLIDRALPLPDGYOQTIIISVSSDTINDA 126
149 DVLAINGASVALSDIPMNGVAVRIGIIDGEYVNPTRKEMSSSTLNLVVAGAPRSQ 208
127 DVLGIGIASAALIMSDIPFANPSEYRVRNGLEFIYVDPDINELAQSMDCIGGT-EDT 185
209 IVMLEASAEMLIQDFCHAIKVGKYYTQIIQGIQOLVKETGVTKRPQKLTFS--PE 265
186 ICMLEGEKKEISBAEMDAIKFG---HDAIKKICALQRELAAEVAKRFRFSPTVADE 241
266 IVKTKHKLAMERLVAVFTDYEDHKYSDAEVANKIRLDEEOLKEKP-----EA 314
242 LVNVEEHCSAEKAL----AYTLAKEERAEKTKAIYOTIRKTLTFTDRVGPDOIEA 297
315 DP-----YEIESFNVAVKEVRSIVLNEYKRCODRDLTSLRNVCSEVDMFKTHGS 366
298 DPTAFACINEMIEECIHAVEKVRHMTLDDKRLDKRLTEQVRPSTIEGLIPRAHGS 357
367 ALFORGQTOVLCTVTPSLSEGISD-QVITAINGIKDKNFMHYEPFPAVTEIGKVTG 425
358 ALPFRGETQALVTITL-----GTKKDAQSDTLTDDKOKRMLHYNPFPSVGEIRGCG 412
426 LNRRELHGALAEALYVIP--RDFPTTIVTSSEVLESNGSSSSMASACGSLALMDSGV 483
413 AGREIGHGNLAERAIKVMPSSEQEPFYTVRLVSDILESNGSSSMASACGSLALMOGI 472
484 PISAVAVGVALGVTKDPKEGIEDYRLLDIIGIEYNGMDPKLAGTKGITALQAD 543
473 PLKRPVSGIANGLKEGD-----RYAVLSDIILNEDHLDMDPKVAGTIDGITALQAD 525
544 IKLPGIPIKIVMEAIQOASVAKKEILQIMNTISKPRASREKNGVAVETVQVPLSKAPF 603
526 IKIDGLVNHILETALQARKGRHLIDVMAAIPESRADIKYAPRLTTIGIPYDALGMV 585
604 VGPBGVNLKLOAETGVTISQVDEBTSVFPPTSVMHEARDFTEICKDQDEQLEFGA 663

DB 586 IKGGEITRISITEETGAENINDDGVTYTIACSSPEATKAAVETIKTLV-----SKPEVGT 640
OY 664 VYATITEIRD-TGVWVKLYPNMTAVLHNTQLDNERL 700
DB 641 IYMKAVDIDDELGAFFPELPK-TDGLVHISEIRARENI 677

RESULT 15
099U8 PRELIMINARY; PRT; 698 AA.

AC 099U8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Polytubonucleotide nucleocidytransferase.
GN NP4 OR SAV1274 OR SA1117.
OS *Staphylococcus aureus* (strain Mu50 / ATCC 700699), and
OS *Staphylococcus aureus* (strain N315).
OC Bacteria; Firmicutes; Bacillales; *Staphylococcus*.
OX NCBI_TaxID=158878, 158879;
RN [1]

RP SEQUENCE FROM N.A.
RC SPECIES=*S.aureus* (strain Mu50), and *S.aureus* (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hoshoyama A.,
RA Mutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki U.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant *Staphylococcus*
RT *aureus*,"
RL Lancet 357:1225-1240(2001).
DR EMBL: AP003361; BAB57436.1; -.
DR EMBL: AP003363; BAB42369.1; -.
DR HSSP: P05055; ISRO.
DR InterPro: IPR001247; 3_ExoRNase.
DR InterPro: IPR004087; KH_dom.
DR InterPro: IPR004088; KH_type_1.
DR InterPro: IPR003029; SL.
DR Pfam: PF00013; KH; 1.
DR Pfam: PF03726; PNPase; 1.
DR Pfam: PF01138; RNase_PH; 2.
DR Pfam: PF03725; RNase_PH_C; 2.
DR Pfam: PF00575; SL; 1.
DR SMART: SM00322; KH; 1.
DR SMART: SM00316; SL; 1.
DR PROSITE: PS50084; KH_TYPE_1; 1.
DR PROSITE: PS50126; SL; 1.
DR Transferrase; Complete proteome.
SQ SEQUENCE 698 AA; 77361 MW; 6EABFB86EDPFB2E CRC64;

Query March 33.5%; Score 1190.5; DB 16; Length 698;
Best Local Similarity 39.1%; Pred. No. 1.3e-66;
Matches 273; Conservative 118; Mismatches 249; Indels 59; Gaps 12;

24 SAGSRAVAVDLGRKKEISSGKLAFADGSANVOSGDTAVMTAVSTKPSQFMFLV 83
2 SQEKVFTKTMAGSLTLETGQAKQANGAVLVRYGTVLSTVATSKPEPDGFFPLTV 61
OY 84 DYKQAAAAGRIPTNVLREVGTSDEKELTSRIIDRSIRPLPAGFYDTQVLCNLAVD 143
DB 62 NYEKVYAAAKIPGCFKRGGRGDDATLTARLIDRIIRLFPKGYGADVQIMMVLASD 121
OY 144 GVNEPDVLAINGASVALSDIPMNGVAVRIGIIDGEYVNPTRKEMSSSTLNLVVAG 203
DB 122 PDGSPQMAAIGSSMALSVSDIPFOGIAGVNAGYIDGKYIINPTVEKEVSRDLLEVAG 181
OY 204 APSQIYMLEASAEMLIQDFCHAIKVGKYYTQIIQGIQOLVKETGVTKRTP 256
DB 182 -HDAVNMVEAGASEITEQMLEIFFGHEIORLVDFQOQIVDHIQPVKQ----- 232

Thu Jan 8 13:54:17, 2004

us-09-907-907a-42.rspc

Page 11

QY 257 QKLFPS-----PEIVKTHKLANERLYAVFTYEHDKVSRDEAVNKIRLDTSEQLKE 309
DB 233 ---FIPAERDEALVERVKSJLTERKGLKEVLTPE-----DKQQRDENLDMNK---BEIVNH 281
QY 310 KPPEADP-----YEIESFNNAVKEVPSIVANEYKRCDCGRDLTSLRNVSCEVDMFKTL 363
DB 282 FTDEEDPENELLKEVYAIINELVKEVRLLADEKIRPDGRKPDDEIRPLDSEVGLPRT 341
QY 364 HGSALFQRCQOTVLCVTPEDESLSGKSDQVITAINGIKOKNFMLYHEFPYPATNEIGKY 423
DB 342 HSGGLFTRCOTQALSVLTIGAL-----GDYQIDGLGPREEKRFMHYHFNPNFNSVGETGPV 397
QY 424 TGLNREELGHGALAEKALYVPVIR-RDPPTIRVTSEVLESNGSSMASACGSLALMDS 481
DB 398 RAPGRREIGHGALGERALKYIIPDTADFPYTRIVSEVLESNGSSQASICGSLTALMDA 457
QY 482 GVPISAVAGVALGLVTKTDPEKEIEDRLTDIGIEDYNGMDMPKTAGTKGITALQ 541
DB 458 GVPKAPVAGIAMGLVTRD-----SYTILTDIQMEDALGDMDFKVAAGTEGITALQ 510
QY 542 ADIKLEPIPIKIVMEALIOQASVAKKEILOIMNKTISKPRASREKNGPVVETQVPLSKRA 601
DB 511 MDIKIDGLTREIEELIEELAEQARRGLBIMNMLOTDQPTRELSAYAPKVVMTIKPKXR 570
QY 602 KEVGPBGYNLKKLOAETGVYISQVDEETSVFAPTPSVHMEADPTIEICKDOEOOLEF 661
DB 571 DVIKPGGKINEIIDETGVKLDIEQDGTIFIGAVDQAMINRAEIEIEITR-----BAEV 625
QY 662 GAVYTTATTEIRDGVWKLYPNMTAVLLHNTOLDNERL 700
DB 626 GQTYQATVKRIEKYGAFGVLPFGKDA-LHISQISKNRI 663

Search completed: January 8, 2004, 10:40:18
Job time : 50 secs

